Hierarchical Probabilistic Inference of Cosmic Shear

M. D. Schneider, D. W. Hogg, P. J. Marshall, W. A. Dawson, D. J. Bard, D. Boutigny, D. Lang, J. Meyers

September 23, 2014

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States government. Neither the United States government nor Lawrence Livermore National Security, LLC, nor any of their employees makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States government or Lawrence Livermore National Security, LLC. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States government or Lawrence Livermore National Security, LLC, and shall not be used for advertising or product endorsement purposes.

HIERARCHICAL PROBABILISTIC INFERENCE OF COSMIC SHEAR

Michael D. Schneider[1, 2], David W. Hogg[3], Philip J. Marshall[4],
William A. Dawson[1], Joshua Meyers[5], Deborah J. Bard[4], Dustin Lang[6]
(Dated: Draft May 14, 2015)
LLNL-JRNL-661076

ABSTRACT

Point estimators for the shearing of galaxy images induced by gravitational lensing involve a complex inverse problem in the presence of noise, pixelization, and model uncertainties. We present a probabilistic forward modeling approach to gravitational lensing inference that has the potential to mitigate the biased inferences in most common point estimators and is practical for upcoming lensing surveys. The first part of our statistical framework requires specification of a likelihood function for the pixel data in an imaging survey given parameterized models for the galaxies in the images. We derive the lensing shear posterior by marginalizing over all intrinsic galaxy properties that contribute to the pixel data (i.e., not limited to galaxy ellipticities) and learn the distributions for the intrinsic galaxy properties via hierarchical inference with a suitably flexible conditional probabilitiy distribution specification. We use importance sampling to separate the modeling of small imaging areas from the global shear inference, thereby rendering our algorithm computationally tractable for large surveys. With simple numerical examples we demonstrate the improvements in accuracy from our importance sampling approach, as well as the significance of the conditional distribution specification for the intrinsic galaxy properties when the data are generated from an unknown number of distinct galaxy populations with different morphological characteristics.

Subject headings: gravitational lensing: weak; methods: data analysis; methods: statistical; catalogs; surveys; cosmology: observations

1. INTRODUCTION

All observations to date are consistent with a cosmological model in which mass is clustered on spatial scales ranging from galaxy sizes of order a kiloparsec to hundreds of megaparsecs, corresponding to angular scales on the sky of tens of arcseconds to several degrees. The cosmological mass distribution acts as a *gravitational lens* for all luminous sources, which imparts inhomogeneous image distortions because the mass distribution is clustered. In principle, the coherent lensing distortions of luminous sources can be used to reconstruct the 3D cosmological mass distribution over most of the observable volume of the universe. However, except near the most dense and rare mass peaks, the gravitational lensing shearing of galaxy images is a percent-level effect perturbing unknown at order unity intrinsic galaxy shapes. The 'cosmic shear' of galaxies therefore can only be detected through statistical correlations of large numbers of galaxy images.

Recently, cosmic shear has been used to place competitive constraints on cosmological model parameters (Jee et al. 2013; Kilbinger et al. 2013; Heymans et al. 2013; MacCrann et al. 2014; Kitching et al. 2014; Kilbinger

schneider42@llnl.gov

[1] Lawrence Livermore National Laboratory, Livermore, CA, 94551, USA.
[2] University of California, Davis, Davis, CA 95616, USA.
[3] Center for Cosmology and Particle Physics, New York University, New York, NY, USA.
[4] SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA.
[5] Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, Stanford, CA 94035, USA.
[6] Department of Physics, Carnegie Mellon University, Pittsburgh, PA, 15213, USA.

2014). All analyses of which we are aware use point estimators for the 2D ellipticities of galaxies in a catalog to infer cosmic shear effects. From the estimated galaxy ellipticity components one can compute a two-point angular correlation function that is related to the mass clustering power spectrum in the standard cosmological framework.

There are several drawbacks to existing cosmic shear inference methods that may severely limit the amount of information that can be extracted about the 3D cosmological mass distribution and underlying cosmological model, for a given set of observations (LSST Dark Energy Science Collaboration 2012). Bernstein & Armstrong (2014) (BA14) give a review of the primary challenges in existing methods including low signal-to-noise, large instrumental and observational systematics, finite pixel sampling, uncertainty in the morphologies of galaxies prior to lensing distortions, and biased ellipticity estimators in the presence of pixel noise.

As in BA14 and also Sheldon (2014), Miller et al. (2007, 2013) we consider a probabilistic model for cosmic shear inference that can in principle avoid many of the weaknesses in existing methods. Unlike in BA14 we initially set aside issues of computational feasibility and aim to specify a statistical framework that is a complete description of the data (i.e., including all physical parameters that are needed to completely describe the measured pixel values in a photometric survey). While we see value in simply writing down the complete statistical framework for cosmic shear we were also able to identify statistical sampling and computational algorithms that are likely to make our framework practical for upcoming surveys. We demonstrated a simple implementation

of our framework in the GREAT3[7] community cosmic shear measurement challenge (Mandelbaum et al. 2013, 2014) and will further demonstrate the computational performance of our framework in subsequent papers.

Most of the parameters that are needed to describe the pixel data are uninteresting for cosmology and can be marginalized out; this is an advantage of probabilistic modeling. If we can infer probability distribution functions (PDFs) for the nuisance parameters and marginalize—instead of either asserting distributions or else fixing the nuisance parameters to heuristically chosen or maximum-likelihood values—we expect to obtain better (more accurate and also more conservative) inferences about the cosmological parameters. Adopting this probabilistic approach moves us (relative to other methods) along the bias–variance trade-off towards less biased inferences. Of course, in the context of fully probabilistic inference, without point estimators, there isn't as clear a definition of "bias" and it may be difficult to put the long literature on weak-lensing biases into context here. However, we expect many of the known biases in point estimators to be ameliorated when we permit the freedom to infer and marginalize out nuisances. In detail, as we give more freedom to the model (more flexibility in the nuisance-parameter space), we will move to even less biased (though also probably less precise) inferences, possibly at large computational cost.

Explicit specification or inference of the distributions of all model parameters not only has the potential to reduce biases in the shear inference but also is the only way to guarantee an optimal measurement (in the sense that no other accurate inference method could yield tighter marginal posterior distributions on the cosmological model parameters). Most past cosmic shear analyses have not specified the distributions of galaxy intrinsic properties and observing conditions (e.g., the PSF) that are assumed in their cosmological inferences. But, because intrinsic galaxy properties and observing conditions do describe important features of the data, all cosmic shear analyses must have (at least implicitly) marginalized over some distributions in these parameters. Implicit marginalization over un-specified priors cannot yield an optimal measurement. Said another way, without explicit marginalization of model parameters it is not possible to saturate the Cramér-Rao bound (e.g. Kendall, M. G. & Stuart, A. 1969). In particular, analyses in which measured ellipticities of galaxies are averaged together over sky patches (or something equivalent) have made an implicit (and strong) assumption about the distribution of intrinsic shapes; this assumption (since unexamined) is likely to be sub-optimal at best.

Once a complete statistical model is specified, there are three key practical challenges for probabilistic shear inference,

1. Specification of an interim model describing galaxy morphologies, flux profiles, and spectral energy distributions. Galaxy features are complex and multifaceted because of the evolution of galaxy properties with cosmic time, galaxy mergers, and varying environments. The *observable* features of galax-

ies are further complicated by object blending that may mimic intrinsic features of the galaxy population (Dawson et al. 2014). We require a parametric model for galaxies and a likelihood function that can describe all such variations in galaxy images over the model parameter ranges that are important for describing the data. With an incomplete galaxy parametric model and without propagating all the information about the model in the likelihood function, it is well known that 'model fitting biases' can be catastrophic for cosmic shear (Voigt & Bridle 2010; Melchior et al. 2010; Bernstein 2010; Kacprzak et al. 2014).

2. Specification of the probability distribution for the galaxy model parameters to be marginalized in the shear inference. While the accurate specification of models for galaxy images may sound challenging to the discerning reader, the specification of the joint distribution of *intrinsic* galaxy properties is perhaps even more daunting. A primary aim of this paper is to describe and demonstrate how a hierarchical model can allow meaningful inference of the properties of the galaxy distribution simultaneously with the shear. Without hierarchical inference (i.e., inferring the distributions of nuisance parameters) we are susceptible to an 'overconfidence problem' wherein asserting a prior that is too simple can yield inferences that appear more precise than warranted.

3. Monte Carlo sampling of model parameters from the joint posterior distribution given multi-epoch imaging of a large galaxy survey. Because both the shear and the PSF vary in correlated ways over any given image, the inferences for different galaxies are statistically dependent. In a straightforward probabilistic inference we would need to fit model parameters simultaneously to all galaxy images, which is a formidable computational challenge. Instead we sample model parameters for each galaxy independently and perform a subsequent joint inference using importance sampling given interim posterior samples for each galaxy image.

In Section 2 we outline our framework for the complete cosmic shear inference problem. In Section 2.1 we enumerate the components the framework and in Section 2.2 we outline the key algorithms we use for Monte Carlo sampling of model parameters to render our framework computationally feasible. In Section 3 we focus on implementation details for just the inference of intrinsic galaxy properties, leaving similar details for the PSF and lensing mass inferences to future work. We describe our model for the distribution of intrinsic galaxy properties in Section 3.1 and give numerical examples of shear and ellipticity inferences in Section 3.2. In Section 4 we discuss how to expand the inferences demonstrated for the intrinsic ellipticity distribution to the PSF and lensing mass models and summarize our conclusions.

2. PART 1: DESCRIPTION OF THE STATISTICAL FRAMEWORK

2.1. *Three conditionally independent branches*

Parameter	Description
$\boldsymbol{\theta}$	Cosmological parameters
Ψ^{IC}	Initial conditons for the 3D gravitational potential
Ψ^{LT}	Late-time 3D gravitational potential
ψ_s	2D lens potential (given source photo-z bin s)
A_s	Parameters for the line-of-sight source distribution
$\Pi_{n_s,i}$	PSF for galaxy n_s observed in epoch i
Ω_i	Observing conditions in epoch i
$\{\omega_{n_s}\}$	Galaxy model parameters; $n_s = 1, \ldots, n_{\mathrm{gal},s}$
$\{\alpha_{n_s}\}$	Parameters for the distribution of $\{\omega_{n_s}\}$
$\{\xi_{n_s}\}$	Scaling parameters for $\{\omega_{n_s}\}$
m, τ	Hyperprior parameters for $\{\xi_{n_s}\}$
$\mathcal{A}$	Hyperparameter for $\{\alpha_{n_s}\}$ classifications
$\{\mathbf{d}_{n_s,i}\}$	Pixel data for galaxies $n_s = 1, \ldots, n_{\mathrm{gal},s}$ in epoch i
$G_0\|a_\eta$	Prior specification for $\{\alpha_{n_s}\}$
s	Source sample (e.g., photo-z bin)
W	Survey window function
$\mathbf{d}_{\mathrm{anc},i}$	Ancillary data for PSF in epoch i
p	Prior params. for observing conditions
a	Prior params. for $\mathcal{A}$
$\sigma_{\mathrm{pix},i}$	Pixel noise r.m.s. in epoch i
I	Model selection assumptions

We begin this section by enumerating the variables and dependencies of a complete statistical framework for shear inference. In subsequent sections we demonstrate the inference of a subset of the variables in our framework with numerical models. We defer a more detailed examination of some aspects of the statistical framework to later publications. But we believe it is useful at this stage to list all contributions to the cosmic shear inference problem given the considerable literature on the subject that has not yet converged on a unified statistical picture.

Under a limited set of assumptions described below, the statistical framework for cosmic shear separates into three conditionally independent branches: 1) cosmological lensing mass density, 2) observational point-spread functions, 3) intrinsic galaxy properties.

In Appendix F we describe the computational implementation of the framework in this Section.

2.1.1. *Lensing mass distribution*

We start by specifying the parameters $\boldsymbol{\theta}$ of the cosmological model (top right of Figure 1), sampled from a prior distribution $\Pr(\boldsymbol{\theta})$ given all past cosmological experiments. Although not explicitly implemented in this paper, we expect $\boldsymbol{\theta}$ to include parameters such as the mean mass density Ω_m and the r.m.s. of mass density fluctuations σ_8 that primarily determine the amplitude of the cosmic shear correlation function.

From these parameters, we can predict the 3D cosmological mass distribution, described by the 3D gravitational lensing potential Ψ. Assuming, for example, Gaussian initial conditions Ψ^{IC} for the 3D mass density in the early universe and gravitational evolution according to General Relativity, the probability distribution for the late-time 3D mass density Ψ^{LT} responsible for gravitational lensing of galaxies depends only on the cosmological parameters and the initial conditions, $\Pr(\Psi^{\mathrm{LT}}|\boldsymbol{\theta}, \Psi^{\mathrm{IC}})$. We discuss possibilities for inferring
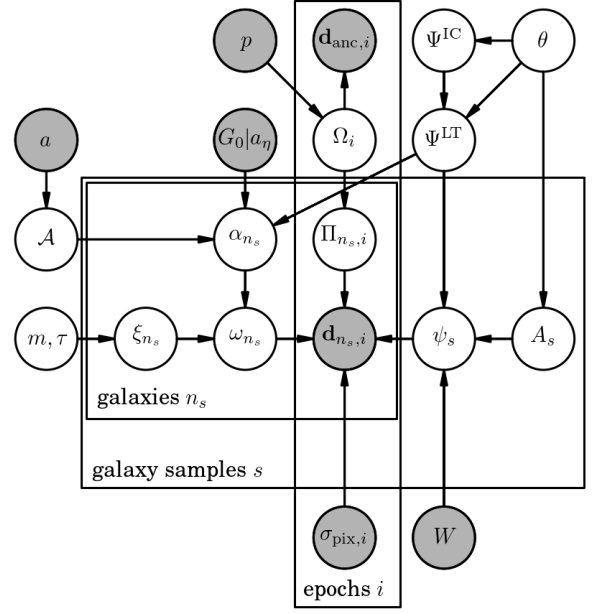

FIG. 1.— Probabilistic graphical model for our complete framework for shear inference. Arrows indicate statistical dependencies. Grey shading indicates quantities that are not sampled. See Table 1 for descriptions of the parameter symbols.

constraints on Ψ^{IC} in Section 2.2.1.0 but otherwise confine our analyses to the inference of gravitational lensing quantities after projecting the 3D gravitational potential Ψ^{LT} over the line-of-sight distribution of lenses in a survey.

We define the 2D lensing potential that describes the shearing (and magnification) of any sample of galaxies in a survey by the following integral of the 3D potential Ψ^{LT} (see, e.g., equation 6.14 of Bartelmann & Schneider 2001) (also section 3.2 of Narayan & Bartelmann 1996),

$$\bar{\psi}_s(\mathbf{x}) \equiv W(\mathbf{x}) \int da \, \Psi^{\mathrm{LT}}(\mathbf{x}, a) K(a; A_s), \qquad (1)$$

where a is the cosmological scale factor in the Freidmann equation, $\mathbf{x}$ are the coordinates in the plane of the sky, $W(\mathbf{x})$ is the angular window function of the survey, the subscript s denotes a particular sample of 'source' galaxies that are lensed by the foreground potential Ψ^{LT}, and K is the lensing kernel including the lensing efficiency based on the distances between observer, lens, and sources, and the integral over the source distribution with parameters A_s (see equation 6.19 of Bartelmann & Schneider 2001).

The parameters A_s can include the uncertainties in the survey redshift distribution (e.g., from photometric redshift errors, see Loh & Spillar 1986; Connolly et al. 1995; Huterer et al. 2006) as well as other astrophysical systematics that are redshift-dependent (such as intrinsic alignments of galaxies, see Catelan et al. 2001; Bridle & King 2007). While both the galaxy redshift distribution and intrinsic alignment correlations might be more properly included in the specification of the properties of the galaxy distribution (Section 2.1.3), we include parameters A_s for the line-of-sight integration of the potentials to connect with more common analysis conventions.

Marginalizing the nuisance parameters in the line-of-sight projection, the 2D lensing potential ψ given the cosmology and survey sample is distributed according to,

$$\Pr(\psi_s|\theta, s, W) \propto \int d\Psi^{\mathrm{LT}}$$
$$\times \int d\Psi^{\mathrm{IC}} \Pr\left(\Psi^{\mathrm{LT}}|\Psi^{\mathrm{IC}}, \boldsymbol{\theta}\right) \Pr\left(\Psi^{\mathrm{IC}}|\boldsymbol{\theta}\right)$$
$$\times \int dA_s \Pr(\psi_s|\Psi^{\mathrm{LT}}, A_s, W)\Pr(A_s|\boldsymbol{\theta}, s, W). \quad (2)$$

For a deterministic cosmological model (as emphasized in Jasche & Wandelt 2013), $\Pr(\psi|\Psi^{\mathrm{LT}}, A, W) \propto \delta_{\mathrm{D}}\left(\psi - \bar{\psi}(\Psi^{\mathrm{LT}}, A, W)\right)$, where δ_{D} is the Dirac delta function. Inserting this expression into Equation (2) yields a trivial solution for the integral over Ψ^{LT}, leaving integrations over the 3D mass density initial conditions Ψ^{IC} and line-of-sight uncertainties A_s to evaluate the marginal distribution of the lens potential ψ given the cosmology $\boldsymbol{\theta}$. Efficient numerical evaluation of Equation (2) is potentially a major computational challenge for our framework, which we will address in subsequent publications.

2.1.2. Point-spread functions

To connect the lens potential to the observable properties of galaxy images, we must also specify the distribution of the point-spread function (PSF) realizations specific to the (time-variable) observing conditions at epoch i given parameters Ω_i such as the seeing, optics alignments, and detector response, $\Pr(\Pi_{n,i}|\Omega_i)$. For given Ω, the PSF $\Pi(\mathbf{H}_n; \Omega)$ can be computed at the focal plane position $\mathbf{H}_n$ for any galaxy n. We model the PSF $\Pi_{n,i}$ as distinct random variables for each galaxy location n and epoch i.

The observing conditions Ω_i in an observation epoch i will not be known perfectly but will be constrained by ancillary data $\mathbf{d}_{\mathrm{anc}}$ and a distribution $\Pr(\Omega_i|p)$ conditioned on time-independent PSF model parameters p (e.g., the allowed distributions of optics distortions or atmospheric turbulence power spectra).

2.1.3. Intrinsic source properties

Finally, the intrinsic properties ω_n of a model for galaxy n (e.g., ellipticity, size, brightness) are described by the distribution

$$\Pr(\omega_n|\alpha, I), \quad (3)$$

where α are parameters in the distribution of the intrinsic galaxy properties and I denotes galaxy modeling assumptions (such as the form of the galaxy morphology parameterization). For example, α could describe the width of the intrinsic (i.e., unlensed) ellipticity distribution of a given galaxy population. And in our contribution to the GREAT3 challenge described in Mandelbaum et al. (2014), the model assumptions I include modeling all galaxies as elliptical Sérsic profiles. It will be important to keep I explicit as we consider possible model-fitting biases in our framework (Voigt & Bridle 2010; Zuntz et al. 2013; Viola et al. 2014; Kacprzak et al. 2014; Bruderer et al. 2015).

In Figure 1, the galaxy properties ω_{n_s} also depend on 'scaling parameters' ξ_{n_s} and, via α_{n_s}, the gravitational potential Ψ^{LT}. We introduce the scaling parameters ξ_{n_s} in Appendix B as a computational convenience in MCMC sampling. The gravitational potential Ψ^{LT} can influence the distribution of galaxy properties because of assembly bias (Wechsler et al. 2002), which is the label in the literature for variations in galaxy (or surrounding dark matter halo) properties with the local surrounding mass density. In a simple description of the dependence of galaxy properties on mass density, it is well established that the amplitude of galaxy clustering correlations depends on galaxy color (e.g., Zehavi et al. 2011). However, because cosmic shear is insensitive to the clustering of source galaxies to first order (Schneider et al. 2002), we do not consider Ψ^{LT} dependencies further in this paper in any model distributions for the intrinsic galaxy properties ω.

2.2. Sampling methods to divide and conquer

2.2.1. Interim sampling of model parameters for individual sources

The pixel data $\mathbf{d}_n$ in the vicinity of a galaxy n enters our framework only through the likelihood function

$$\Pr(\mathbf{d}_n|\omega_n, \psi, \Pi). \quad (4)$$

We can reasonably assume uncorrelated noise in the common case that sky noise (i.e., Poisson noise from sky background) dominates the noise in the pixel values, giving a likelihood function,

$$\ln \Pr(\mathbf{d}_n) = -\frac{1}{2} \sum_{i=1}^{n_{\mathrm{epochs}}} \frac{\left(\mathbf{d}_{n,i} - \bar{\mathbf{d}}\left(\omega_n, \psi, \Pi_i\right)\right)^2}{\sigma_{\mathrm{pix},n,i}^2} + \mathrm{const.}, \quad (5)$$

where $\bar{\mathbf{d}}$ is the model prediction for the pixel values, i indexes multiple observation epochs, and $\sigma_{\mathrm{pix},n,i}$ is the noise r.m.s. per pixel for object n in epoch i. Note the model prediction in the likelihood includes the PSF Π_i for each observation i of a given galaxy n. This makes the model predictions potentially expensive to compute unless the convolution of the PSF with the galaxy model can be done efficiently. We will return to this issue in Appendix A.

In Equation (4) we have assumed that the likelihood function for all sources in an image can be factored into a product of likelihoods for distinct sources. This will not be true in when the pixel noise is correlated for sources that are adjacent on the sky (e.g., when unresolved flux contaminates the pixels of neighboring sources).

The model prediction $\bar{\mathbf{d}}$ for the pixel values of a galaxy image depends on the cosmological mass density ψ through the action of gravitational lensing on the galaxy morphology and flux (in any given bandpass as lensing is achromatic). For simple elliptical models of galaxies, the ellipticity parameter in ω_n is degenerate with the reduced shear $g \equiv \gamma/(1-\kappa)$. However, the statistical distribution of all elements of ω should be isotropic across different galaxies in the sky (although this is not true for small fields, e.g., around a cluster), while the ellipticities of lensed galaxy images are not. We therefore maintain a conceptual distinction between ω_n and ψ in the model for pixel values of a galaxy image $\bar{\mathbf{d}}$ even though this distinction is not computationally meaningful for the simplest models of galaxy images. See also the discussion following Equation (15) below.

The first step in our computational algorithm is to draw 'interim samples' of $(\omega_n, \psi, \Pi_{n,i})$ via Markov Chain Monte Carlo (MCMC) from the likelihood in Equation (5) for a single source (or region of sky) identified in all available exposures of a survey. The method of source identification for selecting the pixel values $\mathbf{d}_n$ in Equation (5) is not important as long as our model predictions for the observed pixels allow for observationally truncated source profiles or multiple overlapping sources (as will be needed for blended objects). Allowing for such selection effects may also admit interim sampling of the combined pixel data from different instruments or surveys.

PSF model hierarchy — Note that we infer a distinct PSF realization $\Pi_{n,i}$ for each galaxy n in each observation epoch i, which are related across galaxies in a given epoch by the common observing conditions Ω_i. The PSFs in different epochs are further related by another level of hierarchy with a common PSF model with parameters p. So, we can also directly marginalize the likelihood over the PSF realizations,

$$
\begin{aligned}
\Pr(\mathbf{d}|\{\omega_n\}, \psi, p) &= \prod_{i=1}^{n_{\text{epoch}}} \int d\Omega_i \Pr(\Omega_i|p) \\
&\times \prod_{n=1}^{n_{\text{gal}}} \int d\Pi_{n,i} \Pr(\Pi_{n,i}|\Omega_i) \Pr(\mathbf{d}_{n,i}|\omega_n, \psi, \Pi_{n,i}) \\
&= \prod_{i=1}^{n_{\text{epoch}}} \int d\Omega_i \Pr(\Omega_i|p) \left[\prod_{n=1}^{n_{\text{gal}}} \Pr(\mathbf{d}_{n,i}|\omega_n, \psi, \Omega_i) \right] \quad (6) \\
&= \prod_{i=1}^{n_{\text{epoch}}} \Pr(\mathbf{d}_i|\{\omega_n\}, \psi, p) \quad (7) \\
&\neq \prod_{n=1}^{n_{\text{gal}}} \Pr(\mathbf{d}_n|\omega_n, \psi, p) \quad (8)
\end{aligned}
$$

Marginalizing the PSF realizations $\Pi_{n,i}$ over all galaxies n in a given epoch i still yields a likelihood function that is separable for separate objects n in Equation (6), conditioned on the observing conditions Ω_i. This shows that more information about the observing conditions Ω_i leads to smaller statistical correlations among the marginal likelihoods for different galaxies independent from the actual PSF realizations. Under our framework therefore, it is the parameters of the distribution from which PSF realizations are drawn that are most important to constrain rather than estimation and interpolation of PSF realizations at different image plane locations.

Lensing mass inference — The intrinsic galaxy properties ω_n and the PSF $\Pi_{n,i}$ can be interim sampled with distinct parameters for each object. We might consider interim sampling uncorrelated shears and magnifications for every object, but in most cases the shear and the intrinsic ellipticity are strongly degenerate for isolated objects (the exception being resolved galaxy images with multiple morphological components that have different responses to the action of an applied shear). To infer a shear or lens potential that varies over the sky, we then need to specify the model correlations between shears at different sky locations and redshifts. This is distinct from conventional algorithms that involve cross-correlating galaxy ellipticity estimators at different sky locations. We instead seek a generative model for the shear correlations. So, the interim sampling of the variable shear or lens potential must happen at a computational step following the interim sampling of individual galaxy and PSF properties.

We can infer variable shear models without a cosmological prediction of the mass density (which can be computationally expensive) by specifying an interim prior for the lens potential ψ_s for a given source distribution. The only requirement is that the mathematical specification of the interim prior for ψ_s allows for spatial correlations that encapsulate the possible cosmological interpretations. For example, even though the late-time lens potential is not Gaussian distributed, we might specify a multi-variate Gaussian prior on ψ for the purposes of interim sampling. Correlations in the lens potential samples could be captured with a covariance motivated from cosmology,

$$
\langle \psi(\mathbf{x})\psi(\mathbf{x}') \rangle = \int \frac{\ell d\ell}{2\pi} C_\ell(|\mathbf{x} - \mathbf{x}'|) J_0(\ell|\mathbf{x} - \mathbf{x}'|), \quad (9)
$$

where C_ℓ is an angular power spectrum computed, e.g., via Equation (1), and J_0 is the zeroth order Bessel function of the first kind. While Equation (9) is motivated from cosmological theory, we do not connect C_ℓ to a cosmological model when interim sampling ψ. Instead, we use Equation (9) to introduce some degree of spatial correlations in the interim samples of ψ that can increase the efficiency of later sampling steps for constraining cosmology as we describe in the next section.

In practice, we first sample galaxy model parameters without a correction for the applied shear and magnification. Given the galaxy model parameters, we can then reinterpret the parameters under an assumed lens potential model because we know how lensing affects the model for the galaxy light profile. We therefore ignore the lensing parameters in the first step of our inference, including a model for lensing shears and magnifications only when combining inferences from distinct sources as we describe next.

2.2.2. *Hierarchical inference via importance sampling*

It is possible to infer model constraints independently for each galaxy and then combine these independent inferences in a hierarchical framework using the technique of 'importance sampling' (e.g. Geweke 1989; Wraith et al. 2009; Hogg et al. 2010).

Our goal with the importance sampling algorithm is to evaluate the likelihood marginalized over individual

galaxy intrinsic properties and PSF realizations[8],

$$\Pr(\mathbf{d}|\alpha, \Omega, \psi) \propto \prod_{n=1}^{n_{\mathrm{gal}}} \int d\omega_n \Pr(\omega_n|\alpha)$$

$$\times \prod_{i=1}^{n_{\mathrm{epoch}}} \int d\Pi_{n,i} \Pr(\Pi_{n,i}|\Omega_i)$$

$$\times \Pr(\mathbf{d}_{n,i}|\omega_n, \psi, \Pi_{n,i}), \quad (10)$$

where

$$\mathbf{d} \equiv \left\{ \{\mathbf{d}_{n,i}\}_{i=1}^{n_{\mathrm{epoch}}} \right\}_{n=1}^{n_{\mathrm{gal}}}, \quad (11)$$

and $\Omega \equiv \{\Omega\}_{i=1}^{n_{\mathrm{epoch}}}$. Using the identity,

$$\int dx \, p(x) f(x) = \int dx \, p(x) g(x) \frac{f(x)}{g(x)}, \quad (12)$$

for arbitrary probability distributions $p(x), f(x)$ and assuming $g(x)$ has nonzero probability mass over the domain where $f(x)$ is nonzero, we can factor Equation (10) as,

$$\Pr(\mathbf{d}|\alpha, \Omega, \psi) \propto \prod_{n=1}^{n_{\mathrm{gal}}} \int d\omega_n \prod_{i=1}^{n_{\mathrm{epoch}}} \int d\Pi_{n,i}$$

$$\times \frac{\Pr(\omega_n|\alpha)}{\Pr(\omega_n|I_0)} \frac{\Pr(\Pi_{n,i}|\Omega_i)}{\Pr(\Pi_{n,i}|I_0)}$$

$$\times \Pr(\mathbf{d}_{n,i}|\omega_n, \psi, \Pi_{n,i}) \Pr(\omega_n|I_0) \Pr(\Pi_{n,i}|I_0). \quad (13)$$

The final line of Equation (13) defines the 'interim posterior' from which we draw samples when we analyze each galaxy independently (following Hogg et al. 2010). We refer to $\Pr(\omega_n|I_0)$ and $\Pr(\Pi_i|I_0)$ as 'interim priors'. The interim priors can be chosen to make computations convenient. We have found flat or broad Gaussian distributions to be sufficient interim prior specifications.

We use the following algorithm to combine samples from the interim posteriors for each galaxy into a hierarchical inference of the global marginal posterior. For each object n we generate K samples $\left[\omega_{nk}, \{\Pi_{nik}\}_{i=1}^{n_{\mathrm{epoch}}}\right]$, sampled from the interim posterior given pixel data for galaxy n in all epochs,

$$\Pr(\omega_n, \Pi_n|\mathbf{d}_n) = \frac{1}{Z_n}$$

$$\times \Pr(\mathbf{d}_n|\omega_n, \Pi_n) \Pr(\omega_n|I_0) \Pr(\Pi_n|I_0), \quad (14)$$

where Z_n is a normalization that we never need to compute, and I_0 denotes the set of assumptions encoded into the interim priors. Once we have this K-element interim sampling for every galaxy n we can build importance-sampling approximations to various other marginalized likelihoods. Specifically, for the integral in Equation (13),

the marginalized likelihood can be approximated by

$$\Pr(\mathbf{d}_n|\alpha, \Omega, \psi) \approx \frac{Z_n}{K} \sum_k \frac{\Pr(\omega_{nk}|\alpha, \psi)}{\Pr(\omega_{nk}|I_0)} \frac{\Pr(\Pi_{nk}|\Omega)}{\Pr(\Pi_{nk}|I_0)}, \quad (15)$$

$$\Pr(\mathbf{d}|\alpha, \Omega, \psi) = \prod_{n=1}^{n_{\mathrm{gal}}} \Pr(\mathbf{d}_n|\alpha, \Omega, \psi). \quad (16)$$

What we have achieved with Equation (15) is the ability to fit galaxy models and PSFs to individual galaxies (via MCMC) and then to combine the fit information from each galaxy into inferences on the distributions of intrinsic galaxy properties α and PSF parameters p. Equation (15) is fast to compute once the K samples for each galaxy are generated. For the final shear inference we will marginalize the parameters α, but with Equation (15) we have already addressed the primary computational bottleneck in modeling images of large numbers of sources in a field.

Although not part of the preceding derivation, we inserted the lensing potential ψ into the list of conditional dependencies in the numerator of Equation (15). This is because given interim posterior samples of galaxy model parameters, we can always reinterpret those model parameters under a different assumption for the shear for any model where we know how to calculate the action of shear on the model galaxy[9].

To evaluate Equation (15) we need K independent samples from the interim posteriors for each galaxy. While this can be slow to generate via MCMC when K is required to be large, we have found in simple tests that even $K = 1$ may be sufficient when assuming a constant shear over a given area of the sky. We use $K = 10$ in the numerical examples in Section 3.

Importance sampling the lens potential— Given interim samples of the lens potential ψ_s we can then infer cosmological parameter constraints from the sampled lens potentials for all galaxy samples in all fields using importance sampling as in Equation (15), but with a new conditional PDF that includes the dependence of the lens potential cosmological parameters $\boldsymbol{\theta}$ and the cosmological initial conditions Ψ^{IC},

$$\Pr(\mathbf{d}_n|\Psi^{\mathrm{IC}}, \theta, W) \propto \frac{1}{N} \sum_{k=1}^{N} \frac{\prod_s \Pr(\hat{\psi}_{s,k}|\Psi^{\mathrm{IC}}, \theta, W)}{\prod_s \Pr(\hat{\psi}_{s,k}|I)}, \quad (17)$$

where, again, we have implicitly marginalized over the line-of-sight distribution error parameters A_s from Equation (2).

Evaluating the numerator on the right-hand side of Equation (17) requires first realizing the 3D mass density perturbations at some early time (e.g., by a traditional initial conditions generator for cosmological N-body simulations). The initial conditions Ψ^{IC} then must be evolved under gravity in an expanding universe (e.g., by running an N-body simulation). Each realization and evolution of Ψ^{IC} is a potentially expensive computation because Ψ^{IC} must cover at least the entire volume probed

---

[8] We may not always want to assume that the PSF realizations for a given galaxy are statistically independent across observation epochs, but it is a useful clarifying assumption here.

[9] This is a distinct concept from estimating the action of shear on the observed pixel values of a galaxy as in BA14.

by the survey in question (tens of gigaparsecs for future surveys) and contain enough small-scale mass resolution to identify model galaxy locations with respect to the cosmological mass density (typically of order tens of kiloparsecs). Thus evaluating Equation (17) for one pair of values of θ and Ψ^{IC} could require running an N-body simulation covering ~ 6 orders of magnitude in dynamic range. Ultimately to infer cosmological parameter constraints we would run an MCMC chain sampling in values of Ψ^{IC} and θ, potentially requiring $\gtrsim 10^4$ expensive cosmological N-body simulations[10].

Inferring cosmological parameter constraints from the procedure just described may sound impractical. However, we are encouraged by noting that,

- Jasche et al. (2014) demonstrated feasible MCMC sampling of 3D density cells of the cosmological mass density by using an analytic perturbative model for the evolution of the mass density from Gaussian initial conditions,

- upcoming surveys such as LSST are already considering in excess of 10^4 expensive cosmological simulations for parameter inferences (Dodelson & Schneider 2013; Taylor et al. 2013),

- several post-processing methods have shown promise for reducing the computation times in generating ensembles of cosmological simulations (Schneider et al. 2011; Angulo & White 2010; Mead & Peacock 2014; Angulo & Hilbert 2015).

The practicality and tradeoffs of these methods remains an area for future research, both with respect to each other and with respect to traditional analyses with two-point function estimators of the mass density.

Equation (17) also helps to understand the utility in separating the galaxy population into different samples s. It is straightforward to derive separate inferences of Ψ^{IC} and $\boldsymbol{\theta}$ for different sub-samples s of the galaxies to test for consistency with respect to unmodeled systematic errors or new physical phenomena not captured by the cosmological model for the mass density evolution. One could similarly infer different distributions of the intrinsic galaxy properties ω for different sub-samples to test for variable model fitting biases and unmodeled redshift evolution in the galaxy population. But, once we have computed the inferences of the 2D lens potential for different sub-samples s, the combined inference of cosmology in Equation (17) is not much more complicated than that for an undivided galaxy sample.

To summarize, by placing an interim prior on ψ that includes spatial correlations over the observed sky area, we stand to gain from a similar computational factorization of the analysis as that achieved with the model for the intrinsic galaxy properties. Most published cosmic shear analyses follow a reverse algorithm where the shear is estimated, turned into two-point correlation estimators and compared with a two-point correlation model. Instead, our framework must start with the two-point correlation model (for a Gaussian density field), realize

the shear field, apply the realized shear to each galaxy model and then compare the pixel model with the data for each galaxy. We consider the computational separation of each of these steps to be a key benefit of the framework we present here.

Finally, one should keep in mind that the relationship between the lens potential defined over the sky and the lensing distortions of measured galaxy images are dependent on the astrometric solution for each exposure. However, we expect this to be a negligible contribution to the shear inference in most cases. Possible centroid shifts of galaxy images should instead be captured by the PSF model.

2.3. *Statistical framework summary*

In Sections 2.1 and 2.2 we introduced a number of parameters with nontrivial interdependencies. Here we collect all the parameters we have introduced and summarize the statistical framework by presenting the factorization of the unmarginalized joint posterior for all sampled model parameters,

$$
\begin{aligned}
\mathrm{Pr}(\boldsymbol{\theta}, \{\psi_s\}, \{\Pi_i, \Omega_i\}, \{\omega_n, \alpha_n, \xi_n\}, \mathcal{A}, \tau, m, a_\eta | \{\mathbf{d}_n\}, \mathcal{X}) \propto \\
\mathrm{Pr}(\boldsymbol{\theta}) \cdot \mathrm{Pr}(\{\psi_s\} | W, \{s\}, \boldsymbol{\theta}) \cdot \mathrm{Pr}(\mathcal{A}|a) \\
\times \left[\prod_{n=1}^{n_{\mathrm{gal}}} \prod_{i=1}^{n_{\mathrm{epoch}}} \mathrm{Pr}(\Pi_{n,i}|\Omega_i, I) \cdot \mathrm{Pr}(\mathbf{d}_{n,i}|\omega_n, \xi_n, \psi_s, \Pi_{n,i}) \right] \\
\times \left[\prod_{n=1}^{n_{\mathrm{gal}}} \mathrm{Pr}(\omega_n | \alpha_n, I) \cdot \mathrm{Pr}(\alpha_n | \mathcal{A}, a_\eta) \cdot \mathrm{Pr}(\xi_n | m, \tau) \right] \\
\times \left[\prod_{i=1}^{n_{\mathrm{epoch}}} \mathrm{Pr}(\Omega_i | p, \mathbf{d}_{\mathrm{anc},i}) \right] \quad (18)
\end{aligned}
$$

where in the first line we collapsed the conditional parameters into,

$$
\mathcal{X} \equiv [\mathbf{d}_{\mathrm{anc}}, \{s\}, W, a, p, I]. \quad (19)
$$

We collect the definitions of all variables in Equation (18) in Table 1 and display the statistical dependencies in Figure 1. The variables $a_\eta, m, \tau, \mathcal{A}$ will be defined in Section 3.

The final line of Equation (18) is the likelihood function for the pixel data $\mathbf{d}_{n,i}$ of galaxy n in observation epoch i, which depends on the intrinsic galaxy properties ω_n, ξ_n, the lensing potential ψ_s for source sample s containing galaxy n, and the PSF $\Pi_{n,i}$. The preceding lines in Equation (18) describe the hierarchical PDFs for the conditional dependencies in the likelihood function, including the important factorizations across distinct galaxies and observation epochs. In Section 4.1 we discuss how marginalization of the conditional dependencies in the likelihood couples the inference from individual galaxies and epochs as determined by the hierarchical parameters that are constant across galaxies and epochs in Equation (18) (e.g., parameters that do not have n or i subscripts). See Appendix F for a description of the computational implementation of this framework.

3. PART 2: IMPLEMENTATION FOR INFERRING INTRINSIC GALAXY PROPERTIES

3.1. *Model for the conditional distribution of intrinsic galaxy properties*

---

[10] The number 10^4 is simply a guess for the required number of MCMC steps to infer marginal constraints on θ from a converged chain.

We choose a non-parametric distribution to describe the galaxy model parameters both because we have little information to guide us on the choice of a parametric distribution and because we need a flexible distribution to minimize any biases in the model parameter inferences (as mentioned in Section 1). We therefore choose a Dirichlet Process (DP) model (e.g. Ferguson 1973; Walker et al. 1999; Neal 2000; Müller & Quintana 2004; Wang & Dunson 2011) for the distribution of intrinsic galaxy properties, with distinct parameters α_n for each galaxy n.

For a given galaxy n we assume the properties ω_n describing the galaxy image are drawn from a (multivariate) Normal distribution with mean 0 and variance parameters α_n. Denote this by, $\omega_n \sim \mathcal{N}(0, \alpha_n)$, where $\mathcal{N}$ indicates the Normal distribution. By including the index n on α_n, we allow for the possibility that every galaxy has properties drawn from a distinct generating distribution. But, we are accustomed to thinking about galaxies as belonging to different types or populations. So a better model might be to select a small number of values for α that characterize different galaxy types. The parameters ω_n for galaxy n would then be drawn from a zero-mean Normal distribution with variance parameters α selected from one of a few possible classes. This is a description of a mixture of Normal distributions for ω.

The DP model for ω can be derived by starting with a mixture of Normal distributions where the number of mixture components is allowed to be arbitrarily large (e.g., Wang & Dunson 2011). We can describe this mathematically with the following hierarchy (Neal 2000, Eqn. 2.1),

$$\omega_n \sim \mathcal{N}(0, \alpha_n), \quad \alpha_n \sim G(\alpha_n | \mathcal{A}), \quad G \sim \mathrm{DP}(\mathcal{A}, G_0). \tag{20}$$

Equation (20) indicates that the properties of galaxy n are drawn from a Normal distribution with variances α_n, as previously described. The variances α_n for the distribution of the properties ω_n of galaxy n are also modeled as random variables, assumed to be drawn from some distribution $G(\alpha_n | \mathcal{A})$ (yet to be specified) with parameter $\mathcal{A}$.

A key feature of the DP model is that $G(\alpha_n | \mathcal{A})$ is a discrete distribution in α_n. If we consider again a mixture of Normal distributions for ω, then $G(\alpha | \mathcal{A})$ would specify the finite and discrete values that α can take in order to describe the parameters of each mixture component. The final statement in Equation (20) says that the discrete support of the distribution $G(\alpha | \mathcal{A})$ is sampled from yet another distribution denoted $\mathrm{DP}(\mathcal{A}, G_0)$ with parameter $\mathcal{A}$ and 'base distribution' G_0. The base distribution describes the continuous range of support from which the discrete support of $G(\alpha | \mathcal{A})$ can be drawn. The parameter $\mathcal{A}$ is called the DP 'precision' or 'clustering' parameter, which we describe next.

For sampling algorithms and to better understand how parameters α_n for distinct galaxies may take identical values it is useful to first marginalize the distribution G in Equation (20) to get the conditional update rule for the parameters α (eq. 2.2 of Neal 2000),

$$\alpha_n | \alpha_1, \ldots, \alpha_{n-1} \sim \frac{1}{n-1+\mathcal{A}} \sum_{\ell=1}^{n-1} \delta_{\mathrm{D}}(\alpha_\ell)$$
$$+ \frac{\mathcal{A}}{n-1+\mathcal{A}} G_0(\cdot), \quad (21)$$

where n indexes galaxies and δ_{D} is the Dirac delta function. Given the discreteness property of the DP there is nonzero probability that draws of α_n will be repeated. Let $\alpha_1^*, \ldots, \alpha_K^*$ be the unique values among $\alpha_1, \ldots, \alpha_{n-1}$, and let N_c be the number of repeats of α_c^* in the 'latent class' c. Then the conditional update distribution can be equivalently written as (eq. 10 of Teh 2010)

$$\alpha_n | \alpha_1, \ldots, \alpha_{n-1} \sim \frac{1}{n-1+\mathcal{A}}$$
$$\left(\sum_{c=1}^{K} N_c \delta_{\mathrm{D}}(\alpha_c^*) + \mathcal{A} G_0(\cdot) \right). \quad (22)$$

This equation is useful because it shows the clustering properties of the DP and the meaning of the parameter $\mathcal{A}$. The value of α_c^* will be assigned to α_n with probability proportional to the number of times α_c^* has been previously drawn, N_c. Note that when $\mathcal{A}$ is large, successive values for α_n are drawn from the base distribution G_0 with high probability, generating a new α_K^* each time. When $\mathcal{A}$ is small, α_n is assigned to one of the previous values of α_c^* with high probability. We will refer to groups of galaxies $\{n\}$ that have equal values of α_k^* as a 'cluster' or 'sub-population', representing potential galaxy type classifications (analogous to the common 'early-' or 'late-type' classifications). Equation (21) shows that larger values of $\mathcal{A}$ will lead to more clusters or categories of galaxy types while bringing $\mathcal{A}$ to zero will force all galaxies to be assigned to the same cluster or type classification. Note that as long as the observations are exchangeable (which is true for sources in an image), the ordering of values in α is arbitrary.

To summarize, we are motivated to use the DP model for the following reasons,

- The DP is analogous to a mixture model when the number of mixture components is learned from the data,

- Samples for galaxy n are naturally informed by those for galaxies $1, \ldots, n-1$, thereby improving the sampling efficiency (for arbitrary ordering of observed galaxies),

- The DP can find unknown clusters in the galaxy parameters, teaching us about galaxy formation and different prospects for shear inference from different data subsets. For example, we show in Section 3.2.2 that knowledge of a galaxy sub-population that is more round than average can improve the shear inference.

With such an algorithm to cluster the features in the galaxy population we are able to simultaneously fit different distributions for each galaxy and exploit the statistical information from a large sample of galaxies. As

we will see in Section 3, the DP model (rather than a less flexible asserted prior distribution) can be an important component of our framework for improving the accuracy and precision of the shear inference.

For later Gibbs sampling, we again consider the mixture model analogy to the DP. We refer to the mixture component that generates a given galaxy image as the latent class for the galaxy. Let c_n denote the latent class assignment to the mixture that generated galaxy n. And let $\alpha_{c_n} \equiv \alpha_c^*$ denote the parameters specifying the distribution of ω_n for latent class c_n. Then, our distribution of the intrinsic galaxy components for galaxy n is

$$\omega_n | c_n, \alpha \sim \Pr\left(\omega_n | \alpha_{c_n}\right). \qquad (23)$$

The latent class assignments c_n are distributed as (Neal 2000, Eqn. 2.3),

$$c_n | \mathbf{p} \sim \mathrm{Discrete}\left(p_1, \ldots, p_M\right), \qquad (24)$$

for some mixing probabilities $\mathbf{p}$,

$$\mathbf{p} \sim \mathrm{Dirichlet}\left(\mathcal{A}/M, \ldots, \mathcal{A}/M\right), \qquad (25)$$

with the clustering parameter $\mathcal{A}$ still the same for all mixture components.

In the infinite mixture limit, Monte Carlo sampling proceeds by only keeping track of those mixture parameters α_c currently associated with an observed galaxy. Neal (2000) showed with his Algorithm 2 that the conditional probabilities for the latent class assignment c_n for observed galaxy n is (Neal 2000, eq. 3.6),

$$\Pr(c_n = c_\ell | c_{-n}, \mathbf{d}_n, \alpha, \mathcal{X}) = b \, N_{-n,c} \Pr(\mathbf{d}_n | \alpha_{c_\ell}, \mathcal{X}),$$
$$\forall \ell \neq n \qquad (26)$$
$$\Pr(c_n \neq c_\ell \forall \ell \neq n | c_{-n}, \mathbf{d}_n, \mathcal{X}) =$$
$$b \, \mathcal{A} \int \Pr(\mathbf{d}_n | \alpha, \mathcal{X}) \, G_0(\alpha) \, d\alpha, \qquad (27)$$

where b is chosen so the probabilities sum to one, $N_{-n,c}$ is the number of galaxies currently assigned to latent class c excluding galaxy n, $\Pr(\mathbf{d}_n | \alpha, \mathcal{X})$ is defined as in Equation (16), and $\mathcal{X}$ denotes the set of conditional dependencies (ψ, Π, I). Note Equations (26) and (27) are similar to Equation (22) with added information from $\Pr(\mathbf{d}_n | \alpha, \mathcal{X})$. We use Equations (26) and (27) to perform Gibbs updates (Tanner 1996) of the latent class assignment variables c_n (see Neal 2000, for more details, but note the mixture probabilites $\mathbf{p}$ are now integrated out). Given an update for $\mathbf{c} \equiv \{c_1, \ldots, c_n\}$, we draw Gibbs updates for the mixture parameters as,

$$\alpha_{c_n} \sim G_0\left(\alpha_{c_n}\right) \prod_{\ell=1}^{N_{c_n}} \Pr(\mathbf{d}_\ell | \alpha_{c_n}, \mathcal{X}), \qquad (28)$$

where N_{c_n} denotes the number of galaxies associated with latent class c_n. That is, the mixture parameters are drawn from the posterior given all the observed galaxies currently associated with mixture class c_n and the prior $G_0(\alpha_{c_n})$.

The ability to use Gibbs sampling for the DP parameters is useful as the number of galaxies (and therefore number of DP parameters) becomes large. But, Gibbs sampling can also pose challenges both for the flexibility of the statistical model and in parallelization of the

Monte Carlo sampling routines. We address the former challenge in the next sub-section.

3.1.1. *Gibbs updates with importance sampling*

Gibbs sampling for the DP hyperparameters is only practical if the integral in Equation (27) can be computed analytically. This typically limits the choice of G_0 to distributions conjugate to the marginal likelihood $\Pr(\mathbf{d}_n | \alpha_{c_n}, \mathcal{X})$ (e.g. Görür & Rasmussen 2010). But, *with importance samples for ω_n, we can perform fast Gibbs updates of the DP hyperparameters with the weaker assumption that G_0 is conjugate to $\Pr(\omega_n | \alpha_{c_n})$.* That is, we are free to specify a DP base distribution without reference to the functional dependence of the model parameters ω_n in the likelihood. This is a significant computational advantage for our problem.

For example, in Section 3.2 we let ω_n be Gaussian distributed with α a variance parameter. Then we require G_0 to be conjugate to the variance of a Gaussian distribution. If, as in Section 3.2, ω_n represents the galaxy ellipticity, then the likelihood function of the pixel data will depend on a nonlinear function of ω_n since the ellipticity is a nonlinear transformation of pixel values. Already in this idealized scenario it is impossible to specify an analytic G_0 that is conjugate to the marginal likelihood $\Pr(\mathbf{d}_n | \alpha_{c_n}, \mathcal{X})$. This illustrates how importance sampling is critical to the success of the DP model specification for the shear inference. In Appendix B we explain how to achieve a wide variety of DP base distributions that are conjugate to the conditional distribution of intrinsic galaxy properties $\Pr(\omega_n | \alpha_{c_n})$.

Applying the importance sampling formula in Equation (15), the marginal likelihood in Equation (26) is

$$\Pr\left(\mathbf{d}_n | \alpha_{c_l}, \mathcal{X}\right) = \frac{Z_n}{N} \sum_{k=1}^{N} \frac{\Pr\left(\omega_{nk} | \alpha_{c_l}\right)}{\Pr\left(\omega_{nk} | I_0\right)}. \qquad (29)$$

Similarly, for the integral in Equation (27),

$$\int \Pr(\mathbf{d}_n | \alpha, \mathcal{X}) \, G_0(\alpha | a_\eta) \, d\alpha = \frac{Z_n}{N} \sum_{k=1}^{N} \frac{\mathrm{Pr}_{\mathrm{marg}}(\omega_{nk} | a_\eta)}{\Pr(\omega_{nk} | I_0)} \qquad (30)$$

where,

$$\mathrm{Pr}_{\mathrm{marg}}(\omega_{nk} | a_\eta) \equiv \int d\alpha_{c_n} \, G_0(\alpha_{c_n} | a) \Pr(\omega_{nk} | \alpha_{c_n}), \quad (31)$$

is the distribution for ω_n marginalized over the conditional PDF parameters α with parameters a specifying the form of the DP base distribution G_0. Again, *the only practical requirement on the DP base distribution G_0 is that we be able to compute Equation (31) analytically or via fast numerical approximations.* Recall that $\Pr(\omega_n | \alpha_{c_n})$ is the distribution of the intrinsic properties of galaxy n given hyperparameters α_{c_n}. We can then restate our requirement on G_0 as a requirement for conjugate priors on the the intrinsic galaxy properties (such as size, flux, ellipticity, and intensity profile slope).

We describe our implementation of a distribution for G_0 in Appendix B. In summary we find that by introducing an auxiliary variable ξ_n such that $\omega_n \equiv \eta_n \xi_n$, we can allow for flexible, but conditionally conjugate distributions for the DP base distribution G_0. By allowing the

conditional PDF on ω_n to be multivariate Gaussian, we can also accommodate correlations in the parameters of a galaxy model.

In Appendix C we describe our sampling model for the DP clustering parameter $\mathcal{A}$ given some expectation for the number of distinct clusters, or categories, of galaxies in the data.

3.2. Example inferences of intrinsic source properties

In this section we use toy models to demonstrate the performance of 1) importance sampling for hierarchical inference of the intrinsic galaxy ellipticity distribution and 2) the DP model as a flexible approach to modeling the distributions of galaxy intrinsic properties. Throughout this section we make the strong assumptions that i) the PSF is known and ii) the shear is constant among all 'observed' galaxies. These assumptions are not realistic but are intended to simplify the demonstration of important numerical features of our statistical framework.

For our toy models we introduce the following notation for ellipticities, e^{int} is intrinsic ellipticity in a galaxy model, $\tilde{e}$ is the galaxy model ellipticity after applying a shear g to the model, $\varepsilon^{\mathrm{obs}}$ is an estimator for ellipticity computed from observed pixel values.

3.2.1. Toy model 1: Benefits of hierarchical inference

Marginalizing the parameters of the intrinsic galaxy ellipticity distribution can reduce biases in the shear inferences in some cases. We illustrate this point with a simple statistical model that we will refer to as 'toy model 1'.

We construct a simplified model for the shear inference with the following assumptions:

1. the pixel data for each galaxy is reduced to 'observed' ellipticity components $\varepsilon^{\mathrm{obs}}$ with observational uncertainties described by uncorrelated Gaussian noise with variance σ_{obs}^2 for all galaxies,

2. the intrinsic galaxy ellipticity e^{int} is generated from a zero-mean Gaussian distribution and is inferred with a similar hyper-prior,

3. we can work in the weak shear regime where $\tilde{e} \approx e^{\mathrm{int}} + g$.

We illustrate the dependencies of toy model 1 in the left panel of Figure 2. This model is simple enough that we can derive the marginal posteriors for both the shear and the intrinsic ellipticity variance analytically. We perform these derivations in the next subsection and then use the results to build some intuition about the benefits of hierarchical inference of distributions of intrinsic galaxy properties in Section 3.2.1.0.

Marginal posteriors for toy model 1 — Because the intrinsic (i.e., unsheared) galaxy shapes are always unknown, we first marginalize over e^{int} in any analysis. For this toy model, marginalizing over the intrinsic ellipticities for each galaxy yields a Gaussian marginal likelihood for the observed ellipticities of each galaxy given a common variance σ_e^2 for the intrinsic galaxy ellipticities,

$$\Pr(\varepsilon^{\mathrm{obs}}|g, \sigma_e^2) \propto$$
$$\int d^2 e^{\mathrm{int}} \mathcal{N}_{\varepsilon^{\mathrm{obs}}}\left(e^{\mathrm{int}} + g, \sigma_{\mathrm{obs}}^2\right) \cdot \mathcal{N}_{e^{\mathrm{int}}}\left(0, \sigma_e^2\right)$$
$$= \mathcal{N}_{\varepsilon^{\mathrm{obs}}}\left(g, \sigma_{\mathrm{obs}}^2 + \sigma_e^2\right). \tag{32}$$

The marginal likelihood for n_{gal} galaxies is just the product of n_{gal} versions of Equation (32),

$$\Pr\left(\{\varepsilon_i^{\mathrm{obs}}\}_{i=1}^{n_{\mathrm{gal}}}|g, \sigma_e^2\right) \propto \prod_{i=1}^{n_{\mathrm{gal}}} \Pr(\varepsilon_i^{\mathrm{obs}}|g, \sigma_e^2)$$
$$= \left(2\pi\sigma^2\right)^{-n_{\mathrm{gal}}/2} \exp\left[-\frac{n_{\mathrm{gal}}}{2\sigma^2}\mathrm{Var}\left\{\varepsilon_i^{\mathrm{obs}}\right\}\right]$$
$$\times \exp\left[-\frac{n_{\mathrm{gal}}}{2\sigma^2}\left|g - \bar{\varepsilon}^{\mathrm{obs}}\right|^2\right], \tag{33}$$

where we defined,

$$\bar{\varepsilon}^{\mathrm{obs}} \equiv \frac{1}{n_{\mathrm{gal}}}\sum_{i=1}^{n_{\mathrm{gal}}} \varepsilon_i^{\mathrm{obs}}, \quad \text{and} \quad \sigma^2 \equiv \sigma_{\mathrm{obs}}^2 + \sigma_e^2. \tag{34}$$

The conditional posterior for the shear g is a Gaussian distribution given n_{gal} galaxy observations and a known (or assumed) value for the intrinsic ellipticity variance σ_e^2,

$$\Pr\left(g|\sigma_e^2, \{\varepsilon_i^{\mathrm{obs}}\}\right) \propto \Pr\left(\{\varepsilon_i^{\mathrm{obs}}\}|g, \sigma_e^2\right) \cdot \mathcal{N}_g\left(0, \sigma_g^2 \mathbb{1}_2\right)$$
$$\propto \mathcal{N}\left(\mu_{cg}, \sigma_{cg}^2 \mathbb{1}_2\right) \tag{35}$$

with,

$$\mu_{cg} \equiv \bar{\varepsilon}^{\mathrm{obs}}\left(1 + \frac{\sigma^2}{n_{\mathrm{gal}}\sigma_g^2}\right)^{-1} \tag{36}$$

$$\sigma_{cg}^2 \equiv \frac{\sigma^2/n_{\mathrm{gal}}}{1 + \frac{\sigma^2}{n_{\mathrm{gal}}\sigma_g^2}}. \tag{37}$$

where σ_g^2 is the variance of the assumed zero-mean 2D multivariate prior on the shear. Equation (35) represents the shear posterior that is realized in any analysis that does not attempt to simultaneously infer the intrinsic ellipticity distribution. It is useful to consider some limiting cases of Equation (35).

Equation (36) suggests defining the composite variable,

$$x \equiv \frac{\sigma^2}{n_{\mathrm{gal}}\sigma_g^2}. \tag{38}$$

Then,

$$\mu_{cg} = \frac{\bar{\varepsilon}^{\mathrm{obs}}}{1+x}, \qquad \sigma_{cg}^2 = \sigma_g^2 \frac{x}{1+x}. \tag{39}$$

So, point estimators for g conditioned on σ_e will become more biased towards zero as x increases (e.g., for small numbers of galaxies, an informative zero-mean shear prior, or a large dispersion in the intrinsic galaxy ellipticities). The variance on the maximum posterior estimator for g becomes arbitrarily small as $x \to 0$ (e.g., as $n_{\mathrm{gal}} \to \infty$) but is bounded from above by the value of σ_g^2 as x becomes large.
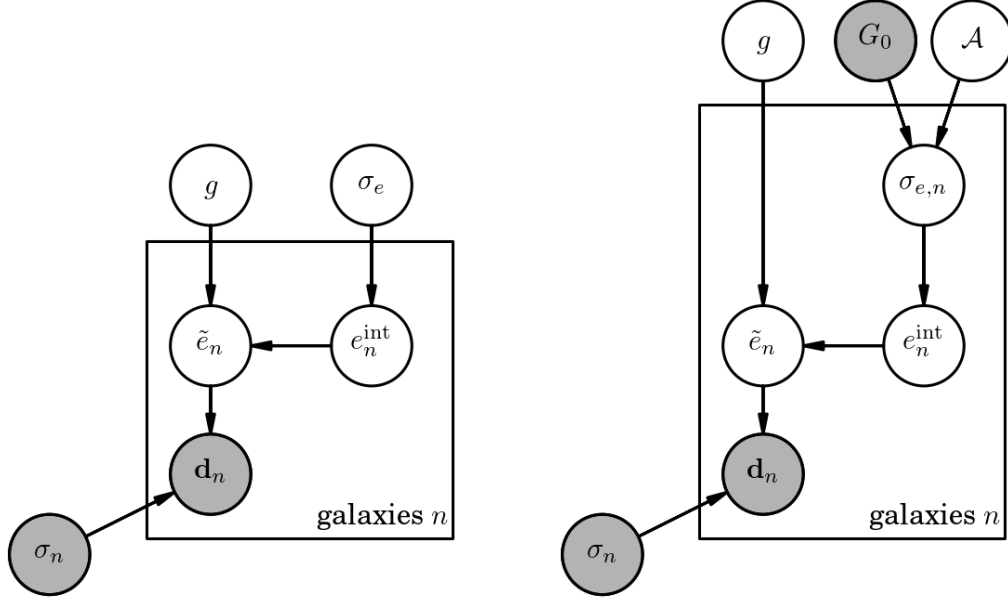
FIG. 2.— Probabilistic graphical models for the two toy models in Section 3.2. Left: Toy model 1 (Section 3.2.1) is set up to illustrate how marginalizing over σ_e reduces biases in the inferred shear g. Right: Toy model 2 (Section 3.2.2) is set up to illustrate how marginalizing over possible distributions for σ_e improves the marginal shear inference in cases where the intrinsic ellipticity is generated from multiple statistical populations.

Marginalizing over σ_e^2 in Equation (32) requires performing the integral,

$$\int_0^\infty dx \, (\sigma_{\text{obs}}^2 + x)^b \exp\left(-\frac{a}{\sigma_{\text{obs}}^2 + x}\right)$$
$$= a^{1-b} \gamma\left(b - 1, \frac{a}{\sigma_{\text{obs}}^2}\right), \quad (40)$$

where $\gamma(b - 1, a/\sigma_{\text{obs}}^2)$ is the lower incomplete Gamma function, and,

$$b \equiv n_{\text{gal}}$$
$$a \equiv n_{\text{gal}} \text{Var}\left\{\varepsilon_i^{\text{obs}}\right\} + n_{\text{gal}} \frac{1}{2}\left|g - \bar{\varepsilon}^{\text{obs}}\right|^2. \quad (41)$$

Marginalizing over the shear gives the following posterior distribution for the intrinsic ellipticity variance,

$$\Pr\left(\sigma_e^2 \middle| \left\{\varepsilon_i^{\text{obs}}\right\}, \sigma_g^2\right)$$
$$\propto \Gamma^{-1}\left(n_{\text{gal}} - 2, n_{\text{gal}} \frac{1}{2}\text{Var}\left\{\varepsilon_i^{\text{obs}}\right\}\right)$$
$$\times \mathcal{N}_{\bar{\varepsilon}^{\text{obs}}}\left(0, \sigma_g^2 + \frac{\sigma^2}{n_{\text{gal}}}\right), \quad (42)$$

where Γ^{-1} is the inverse-Gamma distribution.

We plot the expressions in Equations (35), (40), and (42) for the toy model 1 conditional and marginal posteriors in Figure 3. We have selected parameters for Figure 3, listed in Table 2, that illustrate how a large bias in the shear from the posterior conditioned on σ_e^2 can be alleviated by marginalizing σ_e^2. However, many combinations of parameters in toy model 1 will result in smaller shear biases.

TABLE 2
PARAMETERS FOR TOY MODEL 1 POSTERIOR PLOTS

Parameter	Value
$g_{1,2}$	(0.05, -0.05)
σ_{obs}^2	7.6×10^{-5}
n_{gal}	100
σ_g	0.05
σ_e	0.258
bias in assumed σ_e	0.242

The value of σ_e in Table 2 is derived from a Gaussian fit to the distribution of ellipticity values in the Deep Lens Survey (DLS)[11] (Wittman et al. 2002; Jee et al. 2013) (with shear included, but after correction for the PSF). We assume 100 galaxies in Table 2 both because the shear may not be expected to be constant over a sky area containing larger numbers of galaxies, and because the GNU Scientific Library[12] (Galassi 2009) routine we use to evaluate Equation (40) does not yield numerically stable results for larger n_{gal} values.

The final line of Table 2 indicates the 'bias' in the value of σ_e that is assumed in the conditional posterior in Equation (35), relative to the true value of $\sigma_e = 0.258$. With this bias of 0.242, the assumed value of $\sigma_e^{\text{assumed}} = 0.5$, indicating a value that might be chosen as 'non-informative' in a non-hierarchical analysis. Figure 3 then shows that knowledge of the true distribution of intrinsic ellipticities (via hierarchical inference) can be important in mitigating shear biases and it is not sufficient to assert

---

[11] http://matilda.physics.ucdavis.edu/working/website/index.html
[12] http://www.gnu.org/software/gsl/

a broad intrinsic ellipticity prior.

Prior choices for shear point estimators— Most published cosmic shear analyses rely on point estimators for the shear of each galaxy image (i.e., single-value summary statistics of the shear marginal posterior). We do not advocate the use of point estimators for the shear. But in this subsection we consider the shear posterior mean as a possible point estimator that might be compared with other point estimators in the literature.

The shear posterior mean is often a biased estimator of the shear because the shear posteriors are often non-symmetric. The mean of the shear conditional posterior, Equation (35), is least biased relative to the true shear when σ_g is large and the true value of σ_e is known. The mean of the marginal posterior for the shear, Equation (40), does not require knowledge of σ_e since it has been integrated out and is similarly least biased when σ_g is large. These statements of course depend on our assumption of a zero-mean shear prior.

In Figure 4 we show the bias in the mean of the conditional shear posterior and in the marginal shear posterior as a function of σ_g and the bias in the assumed value of σ_e relative to the value used to generate the data. We can see that the *relative* bias (right panel of Figure 4) in the mean between the conditional and marginal shear posteriors is largest for large σ_g and large assumed σ_e. One might have naïvely guessed that flat priors in shear and intrinsic ellipticity (equivalent in our toy model to large σ_e and σ_g) would be the obvious choice for shear inference in the absence of any other information. But, Figure 4 illustrates that such flat priors will give the most biased shear inference relative to what could be obtained with a hierarchical model for the intrinsic galaxy ellipticities. However, the *absolute* bias in the shear estimator is still non-zero even with the hierachical model, which is one reason we do not advocate the use of point estimators for the shear.

3.2.2. *Toy model 2: DP inference given a bimodal intrinsic ellipticity distribution*

We now construct a numerical toy model to illustrate how the marginal shear inference can be improved by taking advantage of the flexibility of the DP as a conditional PDF for the distribution of intrinsic galaxy properties instead of asserting a simple Gaussian distribution for the galaxy properties. We will refer to the models in this subsection collectively as 'toy model 2'.

As in toy model 1, we assume the pixel data is reduced to observed ellipticities. Unlike in toy model 1 we no longer assume weak shear or a fixed correspondence between the intrinsic ellipticity generating distribution and PDF for inference.

Our objective with toy model 2 is to compare the marginal shear and ellipticity inferences when using either a restrictive or flexible conditional PDF for σ_e^2, irregardless of the generating distribution for the data.

For a restrictive PDF, we specify a uniform prior on a single $\ln(\sigma_e)$ parameter for all galaxies, effectively assuming that all galaxy ellipticities are drawn from a common zero-mean Gaussian distribution. We call this PDF choice our 'Gaussian' model. We use the DP model for σ_e^2 as a flexible hierarchical PDF specification, which we label as our 'DP' model. We describe each model in Ta-

ble 3. Figure 2 shows the relationships of the parameters

TABLE 3
INTRINSIC ELLIPTICITY PDFS FOR TOY MODEL 2.

Model	$\Pr(e_i^{\mathrm{int}})$	$\Pr(\sigma_e^2)$
Gaussian	$N(0, \sigma_e)$	$1/\sigma_e^2$
DP	$N(0, \sigma_{e,i})$	$\mathrm{DP}(\mathcal{A}, G_0)$

in the Gaussian and DP models (left and right panels).

To test and compare the 'Gaussian' and 'DP' models (distinct from the prescription for generating data) we generate two mock data sets, each with 100 galaxies. For each data set we draw samples of e^{int} from a two-component Gaussian mixture,

$$\Pr(e^{\mathrm{int}}) = \lambda \mathcal{N}(0, \sigma_{e,1}) + (1 - \lambda)\mathcal{N}(0, \sigma_{e,2}), \qquad (43)$$

with parameters given in Table 4. The 'unimodal' data

TABLE 4
MOCK DATA PARAMETERS FOR TOY MODEL 2.

Data set	$\sigma_{e,1}$	$\sigma_{e,2}$	mixture fraction λ
Unimodal	0.258	0	1
Bimodal	0.258	0.03	0.7

set has $\lambda = 1$ indicating the e^{int} samples are drawn from a single Gaussian distribution (matching the distribution in toy model 1). For the 'bimodal' data set, we draw e^{int} from a bimodal distribution with 70% of galaxy ellipticities sampled from the Gaussian with width matching that of the unimodal data set, and 30% sampled from a much narrower distribution of intrinsic ellipticity. We describe our detailed procedure for generating mock data sets in Appendix D. Figure 5 shows the observed ellipticity components generated for each mock data set. We motivate the values of $\sigma_{e,1}$ and $\sigma_{e,2}$ in Table 4 by fitting one- or two-component Gaussian mixtures to the distribution of observed ellipticities from the Deep Lens Survey (DLS) (Jee et al. 2013). These ellipticities have been PSF corrected, but include the cosmic shear. We find that a two-component Gaussian mixture is a better fit to the observed ellipticity distribution, giving some motivation for the two-component mixture in our bimodal data set. However, in Table 4 we have artificially increased the fraction of galaxies with a narrow ellipticity distribution over that found in the DLS (where we find a mixture fraction of $\sim 90\%$ for the broader ellipticity distribution).

Figure 6 shows the marginal posteriors for the shear and the intrinsic ellipticity variance as estimated from MCMC samples. We describe the conditional distributions for Gibbs sampling the DP model parameters for toy model 2 in Appendix E. We note here however, that since we have not yet specified an algorithm to determine the DP hyperprior parameters τ and m (see equation B4 and surrounding discussion), we have assigned informative priors for these parameters that are different for the unimodal and bimodal data sets given our knowledge of the generating distributions[13]. An important area for

---

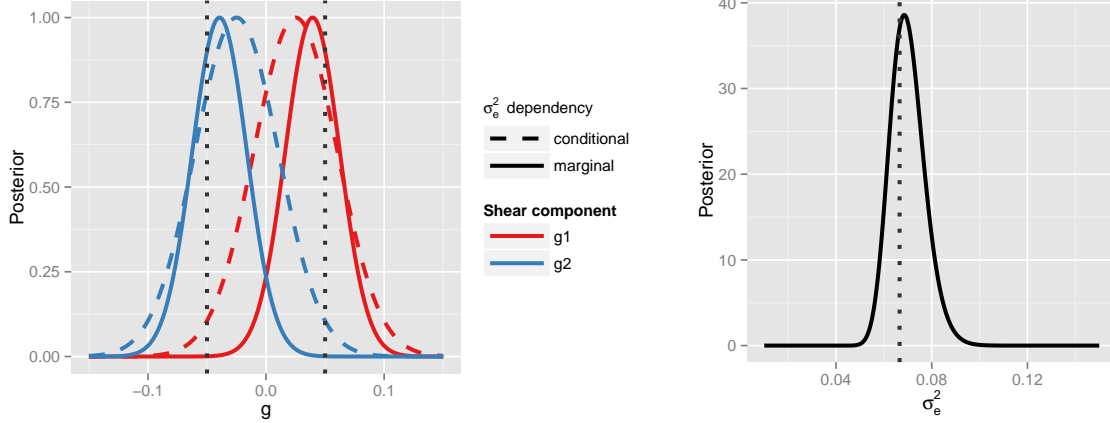[13] Generally we want to define a DP base distribution G_0 with

FIG. 3.— Conditional and marginal posteriors for the shear and intrinsic ellipticity variance in toy model 1. The vertical dotted lines denote the values used to generate the data. See Table 2 for the parameter values used in this figure. Note also the true value of σ_e is 0.258, but the assumed value used to compute the conditional posteriors is 0.5 (indicative of the typical lack of knowledge of the true σ_e value in a cosmic shear analysis).
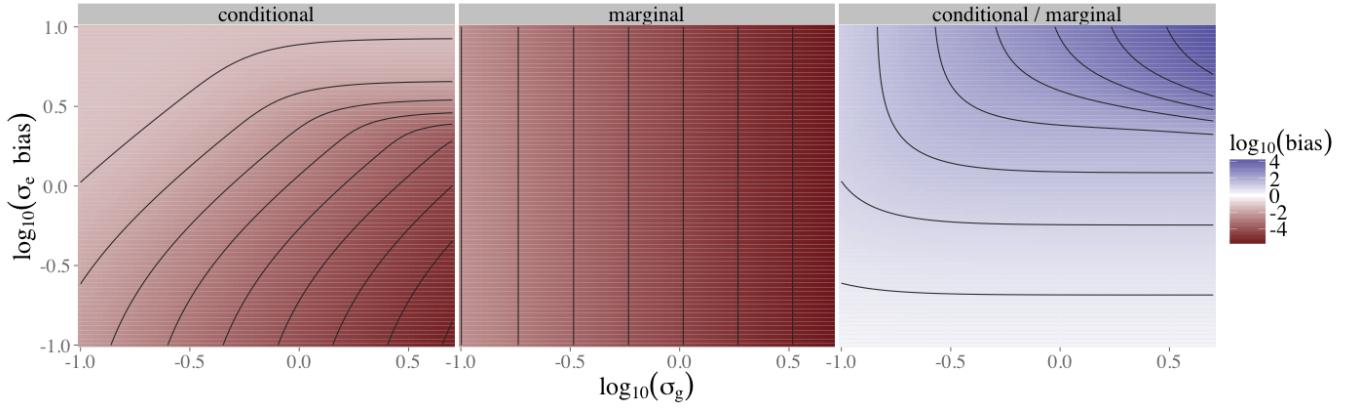

FIG. 4.— Biases in the posterior means for a component of the shear. The left panel shows the posterior mean bias when conditioned on the assumed width of the intrinsic ellipticity distribution, σ_e. The central panel shows that of the posterior marginalized over σ_e (which is insensitive to biases in assumed σ_e). The right panel is the ratio of the shear biases in the left and central panels. The vertical axis shows the logarithm of the additive bias in the assumed σ_e relative to the true value (so a perfect guess for the true width of the intrinsic ellipticity distribution corresponds to $-\infty$ on the vertical axis). The horizontal axis shows the logarithm of the width of the zero-mean prior for the shear components. For all assumed values of σ_e and σ_g, marginalizing over σ_e always gives less biased shear posterior means (comparing the central to the left panel). In addition, assuming broad priors (i.e., weakly informative) for the intrinsic ellipticity and the shear gives the most biased shear inferences if one does not marginalize over the intrinsic ellipticity distribution width (right panel).
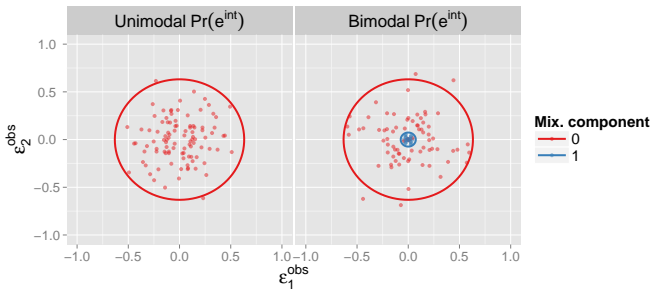

FIG. 5.— Mock data for 100 simulated galaxy ellipticities for toy model 2. The intrinsic ellipticity components are generated from two-component Gaussian mixture models with parameters given in Table 4.

future work will be to dynamically set or sample in the hyperprior parameters for the DP base distribution.

larger probability mass around zero when we expect fewer distinct sub-populations in the data. This is non-intuitive given the variable expansion. $\sigma_e \equiv |\xi|\sigma_\eta$ with $\sigma_\eta \sim DP(\mathcal{A}, G_0)$.

The horizontal black lines in the top panel of Figure 6 indicate the uncertainty on the naïve shear estimator obtained by averaging the observed ellipticities of the galaxies (i.e., the horizontal line gives $\mathrm{sd}\left(\left\{\varepsilon_i^{\mathrm{obs}}\right\}_{i=1}^{n_{\mathrm{gal}}}\right)/\sqrt{n_{\mathrm{gal}}}$). These naïve shear estimator lines are positioned at the height of the 1-σ width of the posterior densities for each model.

TABLE 5
RATIOS OF MEAN MARGINAL SHEAR BIASES USING THE DP OR GAUSSIAN MODELS TO MARGINALIZE THE INTRINSIC ELLIPTICITIES FOR TOY MODEL 2.

Data set	DP model / Gaussian model (g_1, g_2)
Unimodal	(1.26, 1.23)
Bimodal	(0.46, 0.12)

For the marginal shear posteriors with the unimodal data set (top row of the top panel in Figure 6), both the
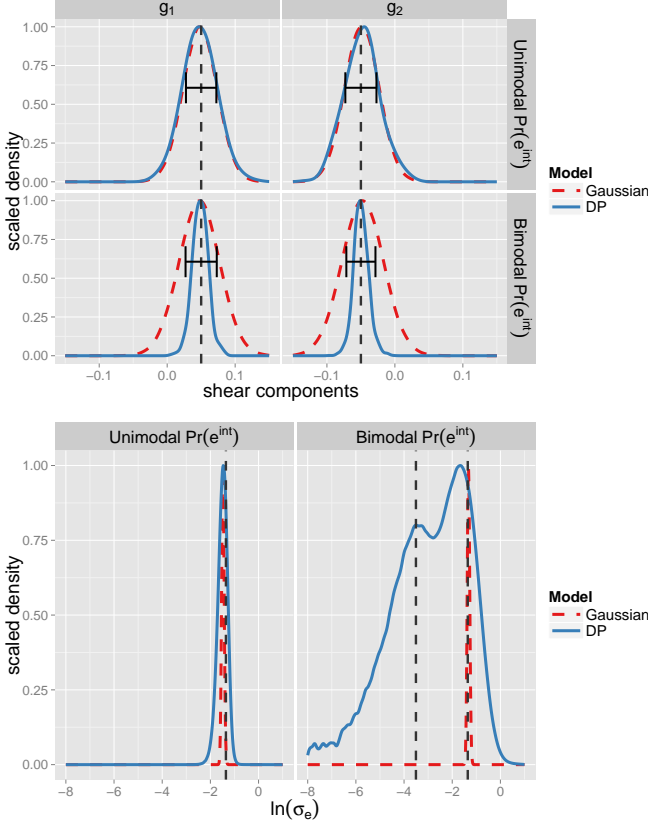
FIG. 6.— Top: Marginal posteriors for the shear components under the four toy model 2 scenarios. The horizontal black lines with error bars indicate the mean and uncertainty on the shear estimator obtained from averaging the observed ellipticity components of all galaxies. Bottom: Marginal posteriors for the logarithm of the standard deviation parameter in the intrinsic ellipticity distribution. For the DP model, we compute marginal densities from the concatenation of MCMC samples for $\sigma_{e,i}$ with $i = 1, \ldots, n_{\text{gal}}$. The vertical dashed lines in all panels show the 'true' values corresponding to the circles in Figure 5.

Gaussian and DP models yield similar results, which are similar as well to results obtained using a simple average of the ellipticities to estimate the shear.

However, the shear marginals for the bimodal data set (lower row of the top panel of Figure 6) show large differences between the Gaussian and DP models. The Gaussian model yields shear marginal distributions similar to those obtained with the unimodal data set, while the DP model yields shear marginals that are more than twice as precise as those assuming the Gaussian model (and are narrower than the shear errors obtained from averaging ellipticities). In Table 5 we give the ratio of the biases in the mean g_1, g_2 from the DP model to that from the Gaussian model for each of the mock data sets. For the bimodal data set, the DP model yields mean shear estimates that are less biased than the Gaussian model, while the reverse is true for the unimodal data set. These results are better understood by next investigating the σ_e marginal distributions in the lower panel of Figure 6.

For the unimodal data set, both models yield accurate inferences of σ_e, although the DP model gives a broader marginal posterior than the Gaussian model. However, for the bimodal data set (right half of the lower panel of Section 3.2, the Gaussian model yields a tight posterior for σ_e that misses the existence of a sub-population

of galaxies generated from a much narrower intrinsic ellipticity distribution. The DP model yields a bimodal marginal posterior for σ_e, indicating some knowledge of the bimodal nature of the generating distribution.

We can now understand the marginal shear posteriors given the bimodal data set as follows. The Gaussian model effectively assumes the entire galaxy population is generated from a broad intrinsic ellipticity distribution, which limits the precision with which the shear can be measured. The widths of the the Gaussian model shear marginals given the unimodal and bimodal data sets are similar because the inference assumes similar width ellipticity generating distributions. However, because the DP model accurately infers the existence of a galaxy subpopulation with small intrinsic ellipticities, this information can be used to more precisely infer the shear for the whole population.
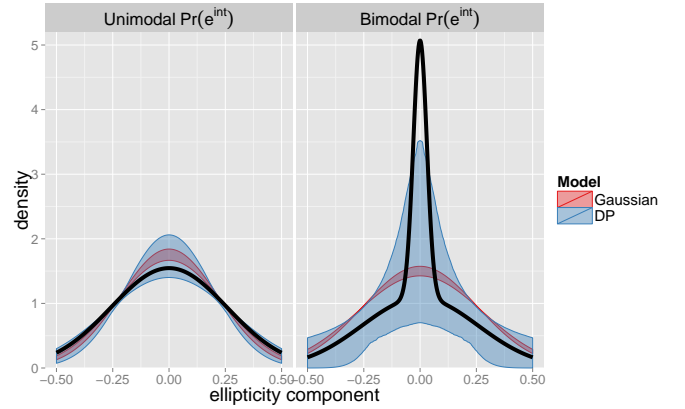

FIG. 7.— Realizations of the intrinsic ellipticity distribution given marginal posterior samples of σ_e in toy model 2. The black lines show the generating distributions for mock data scenarios 'data1' and 'data2'. The shaded bands indicate the 68% credible intervals on the realized ellipticity probability densities. In both mock data scenarios, the credible interval for 'model1', which assumes σ_e is the same for all galaxies, excludes the generating distributions.

Figure 7 shows another way to visualize and interpret the inferences of the intrinsic ellipticity distributions with both models. The solid lines in Figure 7 show the generating distributions of e^{int} for the unimodal and bimodal data sets. The shaded bands show the 68% credible intervals for the inferred e^{int} distributions given marginal posterior samples of σ_e from both models. While the shaded band for the Gaussian model is not dissimilar from the generating distribution for the unimodal data set, it is a poor approximation to the generating distribution for the bimodal data set. We should therefore expect larger shear errors and biases when using the Gaussian model to analyze the bimodal data set than when using a more flexible ellipticity conditional PDF as in the DP model.

In the case of the bimodal generating distribution, we might obtain more precise and accurate marginal shear inferences from a model for the intrinsic ellipticity distribution that is still unimodal, but calibrated from a subsample of the data to partially include information about a rounder subpopulation of galaxies with a different response to shear. The real power of the DP, however, is that it not only helps to infer the underlying distribution of intrinsic ellipticities from the data but that it

also enables assigning different galaxies to different latent classes. Thus, it is capable of determining which galaxies to weight more in the shear measurement (i.e., those that belong to classes with smaller intrinsic ellipticity variance). The DP inference thereby results in not only a less biased shear inference but also a more precise shear inference.
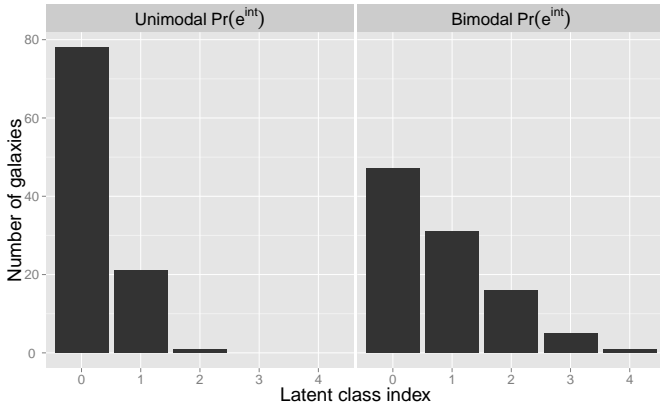

FIG. 8.— The number of galaxies assigned to each of the DP latent classes for a single sample from the MCMC chains for the 'unimodal' and 'bimodal' mock data scenarios in toy model 2. More galaxies assigned to latent classes with larger indices (starting from 0) indicate more distinct clusters of galaxy ellipticities are favored by the DP model.

We can also draw some insight about the galaxy population from the DP sampling in the DP model. The two peaks in the σ_e marginal distribution in Figure 6 indicate the presence of two populations of galaxies with different intrinsic ellipticity distributions in the bimodal data set. We can draw a similar conclusion from the samples of the latent class selectors in the DP model as defined in Equation (24). Figure 8 shows the distributions of latent class indices (given integer labels increasing from 0) for a single sample from the the DP model MCMC chains. We see that the posterior samples given the unimodal data set indeed favor a single generating distribution while those for the bimodal data set favor more than one generating distribution. The shear inference in all cases is marginalized over these latent class assignments.

4. DISCUSSION AND CONCLUSIONS

4.1. *Hierarchical inference for PSFs and shear*

With the toy models in Section 3 we demonstrated how to simultaneously infer a constant shear and the intrinsic galaxy ellipticities. We generated interim posterior samples for each galaxy independently and then combined the inferences from each galaxy using importance sampling and a DP model for the intrinsic ellipticity distribution. By specifying a common hierarchical conditional PDF on the parameters of the intrinsic ellipticity distribution for the galaxy population, we statistically correlated, and thereby improved, the marginal shear inference given all galaxies. A key feature of our model for the intrinsic ellipticity PDF is the assignment of galaxies into latent classes based on the inferred intrinsic ellipticity variance. This allows an improved sensitivity to the action of shear on galaxies with different intrinsic ellipticity magnitudes.

Extending the statistical model to that presented in Equation (18) to include inference of the PSF and shear that is allowed to vary over the sky will have similar effects in correlating the inferences from each galaxy at a hierarchical level. As an aside, the hierarchical model that imposes parameters common to all galaxies in the PDF on, e.g., the intrinsic ellipticity is what allows us to perform interim posterior sampling on individual galaxies as in Section 2.2.2. Similar conclusions hold for both the marginalization of the intrinsic galaxy properties ω and the lensing potential ψ.

The shear is correlated across sky positions because of the common gravitational potentials that distort the light from different galaxies. Ideally, then, we would like to infer the 3D matter potential Ψ from the combination of galaxy shears and redshift information and infer cosmology from the statistics of the 3D mass distribution. While the initial conditions for the 3D mass distributions can be modeled as Gaussian distributed to good approximation, the late-time mass distribution is non-Gaussian and includes significant correlations between Fourier modes of the density field. It is not obvious how to specify a conditional PDF on the 3D mass distribution or how to feasibly sample model parameters (see, e.g., Schneider et al. 2011, for an approach to using cosmological N-body simulations for this purpose). Jasche & Wandelt (2013) and Jasche et al. (2014) have presented reconstructions of the 3D mass density and cosmological initial conditions from galaxy clustering inferred in a spectroscopic survey that might become useful for shear inference as well in the future.

If we pursue a description of the mass density in terms of the realization of the cosmological initial conditions as in Jasche et al. (2014), then we would infer cosmological parameter constraints by marginalizing the initial conditions Ψ^{IC}. The lens potential realizations for different source distributions s all depend on the same initial conditions realization Ψ^{IC} evolved under the same cosmology $\boldsymbol{\theta}$. So, when we marginalize the initial conditions realization Ψ^{IC} the distributions of ψ_s are no longer separable for different s. This is consistent with the common intuition that the lens potentials for, e.g., different photo-z bins, should be correlated by common lensing mass along the lines of sight. This is also why the probability distribution of $\{\psi_s\}$ for all source distributions s does not factor over s in Equation (18).

4.2. *No more point estimators for galaxy catalogs*

In traditional algorithms, point estimators for the ellipticity of an individual galaxy can be computed from a 'postage stamp' image of the galaxy isolated from the rest of the survey imaging. The ellipticity estimates from individual galaxies can then be averaged over local regions of the sky to estimate the shear.

Our likelihood in Equation (5) also applies to galaxy 'postage stamps'. But, there is no statistically meaningful method to combine the information from individual postage stamps unless we draw interim posterior samples from each postage stamp and use methods such as in Section 2.2.2. We come to an important conclusion then that we should dispense with computing point estimators from postage stamp images of galaxies for the purposes of later model inferences (this point was also made previously in Brewer et al. 2013; Hogg 2012).

4.3. *Conclusions*

We have described a hierarchical statistical framework to infer cosmic shear from galaxy imaging surveys. By explicit specification and marginalization of the statistical models for intrinsic (i.e., unlensed) galaxy properties and the PSF, we may be able to improve both the accuracy and precision of the shear inference over previously published algorithms. With a simple toy model in Section 3 we showed promising improvements in the precision of the marginal shear inference when galaxy ellipticities are generated from a bimodal distribution (intended to mimic the different morphological classes of observed galaxies).

We assume a likelihood function for the imaging pixel data. Because we estimate ellipticities and shears as galaxy model parameters rather than reduced statistics of the data, defining a likelihood for pixel values poses no algorithmic challenges for our framework over using reduced statistics. With an importance sampling algorithm we can infer model parameter constraints for independent galaxy images and combine these inferences into a consistent shear inference over many galaxies. We are only able to perform this importance sampling mathematically by means of a hierarchical model that relates the inferences from individual galaxies by means of common parameters in the *distributions* over galaxy properties and PSFs. Our framework also strongly motivates the use of random samples of posterior model parameters rather than point estimators for constructing source catalogs for cosmic shear analysis.

We also introduced a random process to model the distribution of intrinsic galaxy properties (such as ellipticity and size). Our Dirichlet Process (DP) model can fit a wider range of distributions than any asserted specific functional form. In Section 3.2.2 we showed that biased inference of, e.g., the intrinsic galaxy ellipticity distribution can be another form of 'model fitting bias' in the marginal shear constraints. Our DP model helps in alleviating the bias in the posterior median shear (e.g., Table 5) when the generating distribution is multi-modal, but is even more important in reducing the posterior uncertainty on the shear (Figure 6). When we assume a single Gaussian for the generating distribution of galaxy ellipticities the DP model yields shear inferences that are more biased and no more precise than assuming a Gaussian prior on the ellipticities (the Gaussian model with the unimodal data set in toy model 2 of Section 3.2.2). However, this scenario is a limiting case in that we do not expect the distribution of properties of the galaxy population to be well-described by a unimodal Gaussian distribution, and we do not expect to be able to perfectly match our model distribution to the true generating distribution. The similar marginal shear inferences we obtain with flat or DP ellipticity variance PDFs for the unimodal data set in our toy model is therefore a validation of our DP model distribution, which can still perform well when the generating distribution for galaxy ellipticities is more complex. An important topic to explore further is algorithmic determination of all hyperprior parameters in the DP model.

While we believe the importance sampling and DP algorithms we presented will be useful for the analysis of upcoming large surveys, considerable work remains in implementing variations of these algorithms for large-scale imaging analysis.

Our interim posterior sampling given galaxy pixel data is computationally limited both by galaxy model parameter optimizations and by MCMC sampling. Future developments in fast analytic models for galaxy profiles could reduce the computational requirements at this step. The model in Spergel (2010) is an intriguing example of such model variations worth exploring. We also speculate that conventional model-fitting biases, resulting from galaxy models that cannot describe all important features in the data, may be partially alleviated by a flexible model for the PDF on the galaxy model parameters. Because both a change of galaxy model and a change of conditional PDF on a given galaxy model are changes in the integration measure when marginalizing intrinsic galaxy properties it is possible that a suitable conditional PDF specification can reduce remaining model fitting biases.

Our implemention of DP parameter inference relies on Gibbs sampling, which must proceed sequentially in all model parameters. This sampling can become a computational bottleneck as the number of latent classes increases in the DP (with a corresponding increase in the number of parameters to Gibbs update). Parallelizing the Gibbs updates may become useful in cases where the number of latent classes is large. However, optimizing the computational performance of the DP sampling is an area for future work.

Our framework also offers interesting possibilities to infer the distribution of intrinsic galaxy properties (conditioned on some galaxy modeling assumptions), after marginalizing over the shear and PSF. Again, work remains to remove any ad-hoc assumptions in the DP model to obtain confidence in any inferences about galaxy model parameters that are independent of implicit prior assumptions. That is, we want to only impose explicit prior information about the galaxy properties without biases imposed by modeling limitations.

Finally, we must still demonstrate the PSF inference and marginalization using our framework with pixel-level image simulations. This capability is already part of The Tractor code that we used in Mandelbaum et al. (2014), but we have not propagated PSF uncertainties to the shear inference. We are also pursuing methods to incoporate ancillary information about the PSF (e.g., from site-monitoring data, wavefront sensors, or engineering images) using the framework in this paper.

REFERENCES

Angulo, R. E. & Hilbert, S. 2015, MNRAS, 448, 364 [2.2.2.0]

Angulo, R. E. & White, S. D. M. 2010, MNRAS, 405, 143 [2.2.2.0]

Antoniak, C. E. 1974, The Annals of Statistics, 2, 1152 [C]

Bartelmann, M. & Schneider, P. 2001, Phys. Rep., 340, 291 [2.1.1, 2.1.1]

Bernstein, G. M. 2010, MNRAS, 406, 2793 [1]

Bernstein, G. M. & Armstrong, R. 2014, MNRAS, 438, 1880 [1]

Brewer, B. J., Foreman-Mackey, D., & Hogg, D. W. 2013, AJ, 146, 7 [4.2]

Bridle, S. & King, L. 2007, New Journal of Physics, 9, 444 [2.1.1]

Bruderer, C., Chang, C., Refregier, A., Amara, A., Berge, J., & Gamper, L. 2015, ArXiv e-prints [2.1.3]

Catelan, P., Kamionkowski, M., & Blandford, R. D. 2001, MNRAS, 320, L7 [2.1.1]

Connolly, A. J., Csabai, I., Szalay, A. S., Koo, D. C., Kron, R. G., & Munn, J. A. 1995, AJ, 110, 2655 [2.1.1]

Dawson, W. & Schneider, M. 2014, Complementarity of LSST and WFIRST: Regarding Object Blending, Tech. rep. [F]

Dawson, W. A., Schneider, M. D., Tyson, J. A., & Jee, M. J. 2014, ArXiv e-prints [1, F]

Dodelson, S. & Schneider, M. D. 2013, Phys. Rev. D, 88, 063537 [2.2.2.0]

Dorazio, R. M. 2009, Journal of Statistical Planning and Inference, 139, 3384 [C]

Escobar, M. D. & West, M. 1995, Journal of the American Statistical Association, 90, 577 [C]

Ferguson, T. S. 1973, The Annals of Statistics, 1, 209 [3.1]

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306 [A]

Galassi, M. e. a. 2009, Gnu Scientific Library Reference Manual, 3rd edn. (Network Theory Ltd) [3.2.1.0]

Gelman, A. 2006, Bayesian Analysis, 1, 1 [B]

Geweke, J. 1989, Econometrica, 57, 1317 [2.2.2]

Görür, D. & Rasmussen, C. E. 2010, Journal of Computer Science and Technology, 25, 653 [3.1.1]

Heymans, C., Grocutt, E., Heavens, A., Kilbinger, M., Kitching, T. D., Simpson, F., Benjamin, J., Erben, T., Hildebrandt, H., Hoekstra, H., Mellier, Y., Miller, L., Van Waerbeke, L., Brown, M. L., Coupon, J., Fu, L., Harnois-Déraps, J., Hudson, M. J., Kuijken, K., Rowe, B., Schrabback, T., Semboloni, E., Vafaei, S., & Velander, M. 2013, MNRAS, 432, 2433 [1]

Hogg, D. W. 2012, Hogg's Research Blog [4.2]

Hogg, D. W., Myers, A. D., & Bovy, J. 2010, ApJ, 725, 2166 [2.2.2, 2.2.2]

Huterer, D., Takada, M., Bernstein, G., & Jain, B. 2006, MNRAS, 366, 101 [2.1.1]

Jain, B., Spergel, D., Bean, R., Connolly, A., Dell'antonio, I., Frieman, J., Gawiser, E., Gehrels, N., Gladney, L., Heitmann, K., Helou, G., Hirata, C., Ho, S., Ivezić, Ž., Jarvis, M., Kahn, S., Kalirai, J., Kim, A., Lupton, R., Mandelbaum, R., Marshall, P., Newman, J. A., Perlmutter, S., Postman, M., Rhodes, J., Strauss, M. A., Tyson, J. A., Walkowicz, L., & Wood-Vasey, W. M. 2015, ArXiv e-prints [F]

Jasche, J., Leclercq, F., & Wandelt, B. D. 2014, ArXiv e-prints [2.2.2.0, 4.1]

Jasche, J. & Wandelt, B. D. 2013, MNRAS, 432, 894 [2.1.1, 4.1]

Jee, M. J., Tyson, J. A., Schneider, M. D., Wittman, D., Schmidt, S., & Hilbert, S. 2013, ApJ, 765, 74 [1, 3.2.1.0, 3.2.2]

Kacprzak, T., Bridle, S., Rowe, B., Voigt, L., Zuntz, J., Hirsch, M., & MacCrann, N. 2014, MNRAS, 441, 2528 [1, 2.1.3]

Kendall, M. G. & Stuart, A. 1969, The Advanced Theory of Statistics, Vol. II (Van Nostrand, New York) [1]

Kilbinger, M. 2014, ArXiv e-prints [1]

Kilbinger, M., Fu, L., Heymans, C., Simpson, F., Benjamin, J., Erben, T., Harnois-Déraps, J., Hoekstra, H., Hildebrandt, H., Kitching, T. D., Mellier, Y., Miller, L., Van Waerbeke, L., Benabed, K., Bonnett, C., Coupon, J., Hudson, M. J., Kuijken, K., Rowe, B., Schrabback, T., Semboloni, E., Vafaei, S., & Velander, M. 2013, MNRAS, 430, 2200 [1]

Kitching, T. D., Heavens, A. F., Alsing, J., Erben, T., Heymans, C., Hildebrandt, H., Hoekstra, H., Jaffe, A., Kiessling, A., Mellier, Y., Miller, L., van Waerbeke, L., Benjamin, J., Coupon, J., Fu, L., Hudson, M. J., Kilbinger, M., Kuijken, K., Rowe, B. T. P., Schrabback, T., Semboloni, E., & Velander, M. 2014, MNRAS, 442, 1326 [1]

Loh, E. D. & Spillar, E. J. 1986, ApJ, 303, 154 [2.1.1]

LSST Dark Energy Science Collaboration. 2012, ArXiv e-prints [1]

MacCrann, N., Zuntz, J., Bridle, S., Jain, B., & Becker, M. R. 2014, ArXiv e-prints [1]

Mandelbaum, R., Rowe, B., Armstrong, R., Bard, D., Bertin, E., Bosch, J., Boutigny, D., Courbin, F., Dawson, W. A., Donnarumma, A., Fenech Conti, I., Gavazzi, R., Gentile, M., Gill, M. S. S., Hogg, D. W., Huff, E. M., Jee, M. J., Kacprzak, T., Kilbinger, M., Kuntzer, T., Lang, D., Luo, W., March, M. C., Marshall, P. J., Meyers, J. E., Miller, L., Miyatake, H., Nakajima, R., Ngole Mboula, F. M., Nurbaeva, G., Okura, Y., Paulin-Henriksson, S., Rhodes, J., Schneider, M. D., Shan, H., Sheldon, E. S., Simet, M., Starck, J.-L., Sureau, F., Tewes, M., Zarb Adami, K., Zhang, J., & Zuntz, J. 2014, ArXiv e-prints [1, 2.1.3, 4.3]

Mandelbaum, R., Rowe, B., Bosch, J., Chang, C., Courbin, F., Gill, M., Jarvis, M., Kannawadi, A., Kacprzak, T., Lackner, C., Leauthaud, A., Miyatake, H., Nakajima, R., Rhodes, J., Simet, M., Zuntz, J., Armstrong, B., Bridle, S., Coupon, J., Dietrich, J. P., Gentile, M., Heymans, C., Jurling, A. S., Kent, S. M., Kirkby, D., Margala, D., Massey, R., Melchior, P., Peterson, J., Roodman, A., & Schrabback, T. 2013, ArXiv e-prints [1]

Mead, A. J. & Peacock, J. A. 2014, MNRAS, 440, 1233 [2.2.2.0]

Melchior, P., Böhnert, A., Lombardi, M., & Bartelmann, M. 2010, A&A, 510, A75 [1]

Miller, L., Heymans, C., Kitching, T. D., van Waerbeke, L., Erben, T., Hildebrandt, H., Hoekstra, H., Mellier, Y., Rowe, B. T. P., Coupon, J., Dietrich, J. P., Fu, L., Harnois-Déraps, J., Hudson, M. J., Kilbinger, M., Kuijken, K., Schrabback, T., Semboloni, E., Vafaei, S., & Velander, M. 2013, MNRAS, 429, 2858 [1]

Miller, L., Kitching, T. D., Heymans, C., Heavens, A. F., & van Waerbeke, L. 2007, MNRAS, 382, 315 [1]

Müller, P. & Quintana, F. A. 2004, Statistical Science, 19, 95 [3.1]

Murugiah, S. & Sweeting, T. 2012, Journal of Statistical Planning and Inference, 142, 1947 [B, C, C]

Narayan, R. & Bartelmann, M. 1996, ArXiv Astrophysics e-prints [2.1.1]

Neal, R. M. 2000, Journal of computational and graphical statistics [3.1, 3.1, 3.1, 3.1, 3.1, E]

Schneider, M. D., Cole, S., Frenk, C. S., & Szapudi, I. 2011, The Astrophysical Journal, 737, 11 [2.2.2.0, 4.1]

Schneider, P., van Waerbeke, L., & Mellier, Y. 2002, A&A, 389, 729 [2.1.3]

Sheldon, E. S. 2014, MNRAS, 444, L25 [1]

Soumagnac, M. T., Abdalla, F. B., Lahav, O., Kirk, D., Sevilla, I., Bertin, E., Rowe, B. T. P., Annis, J., Busha, M. T., Da Costa, L. N., Frieman, J. A., Gaztanaga, E., Jarvis, M., Lin, H., Percival, W. J., Santiago, B. X., Sabiu, C. G., Wechsler, R. H., Wolz, L., & Yanny, B. 2013, ArXiv e-prints [F]

Spergel, D. N. 2010, The Astrophysical Journal Supplement, 191, 58 [4.3]

Tanner, Martin, A. 1996 (Springer-Verlag) [3.1]

Taylor, A., Joachimi, B., & Kitching, T. 2013, MNRAS, 432, 1928 [2.2.2.0]

Teh, Y. W. 2010, in Encyclopedia of Machine Learning (Springer) [3.1]

Viola, M., Kitching, T. D., & Joachimi, B. 2014, MNRAS, 439, 1909 [2.1.3]

Voigt, L. M. & Bridle, S. L. 2010, MNRAS, 404, 458 [1, 2.1.3]

Walker, S. G., Damien, P., Laud, P. W., & Smith, A. F. M. 1999, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61, 485 [3.1]

Wang, L. & Dunson, D. B. 2011, Journal of Computational and Graphical Statistics, 20, 196 [3.1]

Wechsler, R. H., Bullock, J. S., Primack, J. R., Kravtsov, A. V., & Dekel, A. 2002, ApJ, 568, 52 [2.1.3]

Wittman, D. M., Tyson, J. A., Dell'Antonio, I. P., Becker, A.,
Margoniner, V., Cohen, J. G., Norman, D., Loomba, D.,
Squires, G., Wilson, G., Stubbs, C. W., Hennawi, J., Spergel,
D. N., Boeshaar, P., Clocchiatti, A., Hamuy, M., Bernstein, G.,
Gonzalez, A., Guhathakurta, P., Hu, W., Seljak, U., &
Zaritsky, D. 2002, in Society of Photo-Optical Instrumentation
Engineers (SPIE) Conference Series, Vol. 4836, Survey and
Other Telescope Technologies and Discoveries, ed. J. A. Tyson
& S. Wolff, 73–82 [3.2.1.0]

Wraith, D., Kilbinger, M., Benabed, K., Cappé, O., Cardoso,
J.-F., Fort, G., Prunet, S., & Robert, C. P. 2009, Phys. Rev. D,
80, 023507 [2.2.2]
Zehavi, I., Zheng, Z., Weinberg, D. H., Blanton, M. R., Bahcall,
N. A., Berlind, A. A., Brinkmann, J., Frieman, J. A., Gunn,
J. E., Lupton, R. H., Nichol, R. C., Percival, W. J., Schneider,
D. P., Skibba, R. A., Strauss, M. A., Tegmark, M., & York,
D. G. 2011, ApJ, 736, 59 [2.1.3]
Zuntz, J., Kacprzak, T., Voigt, L., Hirsch, M., Rowe, B., &
Bridle, S. 2013, MNRAS, 434, 1604 [2.1.3]

APPENDIX

MODELING GALAXY IMAGES

To evaluate the likelihood of the pixel values in Equation (5) we require a computationally efficient code to predict pixel values given a galaxy model that can be sheared, convolved with a given PSF, and depositied on pixels with the addition of noise. THE TRACTOR public code[14] performs exactly these operations. A key feature of THE TRACTOR is that both galaxy flux and PSFs are represented as sums of 2D Gaussian distributions allowing fast analytic convolutions of model galaxy images with a model PSF.

For an isolated galaxy image (or "postage stamp"), we generate posterior samples of the galaxy model parameters ω via MCMC using THE TRACTOR to evaluate the likelihood at each MCMC step. In all tests of our algorithms we have used the public MCMC code[15] emcee (Foreman-Mackey et al. 2013) for inferring galaxy model parameters and PSFs.

The initialization of the galaxy model parameters via numerical optimization can be the least accurate part of the calculation to infer posterior constraints on a galaxy model. We have found that an informed choice of 'interim prior' on the galaxy model can significantly improve the inferences for individual galaxies.

In the next sub-section we describe how to properly incorporate the posterior samples for individual galaxy models with per-galaxy interim prior information into the global statistical inference for a survey of galaxies. We reserve all further discussion of fitting galaxy models to pixel data to a later publication.

CHOOSING A BASE DISTRIBUTION FOR THE INTRINSIC GALAXY PROPERTIES

We can further understand some of the choices in specifying a DP base distribution G_0 by considering the model where

$$\Pr(\omega|\alpha) = \mathcal{N}_\omega(0, \Sigma) \tag{B1}$$

such that

$$\alpha \equiv \Sigma \tag{B2}$$

is the (co)variance of the distribution of intrinsic galaxy properties.

In the case that ω_n is a single parameter so that $\Sigma \equiv \sigma^2$, an obvious conjugate prior for σ^2 is the inverse-Gamma (IG) distribution. However, it is difficult to specify a non-informative prior on σ^2 using the IG distribution when the values of σ^2 are likely to be much less than one. Another choice is the uniform distribution, but this too often leads to biased inference of the variance parameter (Gelman 2006).

Instead we follow Gelman (2006) and impose a more flexible, but still conditionally conjugate, distribution on σ_i^2 by decomposing $\omega_i = |\xi_i| \eta_i$, where ξ is a scaling parameter and $\eta_i \sim \mathcal{N}(0, \sigma_\eta^2)$. We specify distributions $\xi_i \sim \mathcal{N}(m, \tau^{-1})$ and $\sigma_\eta^2 \sim \mathrm{DP}(\mathcal{A}, G_0)$. The standard deviation parameter is $\sigma = |\xi|\sigma_\eta$. The DP base distribution (now for σ_η^2) is,

$$G_0(\sigma_\eta^2) = \Gamma^{-1}(a_\eta, 0.5). \tag{B3}$$

The family of distributions parameterized by a_η, m, τ is quite flexible (Murugiah & Sweeting 2012). To recover a simpler half-Cauchy PDF we take $m = 0$ so that

$$\Pr\left(\sigma^2|\tau, a_\eta\right) \propto \left(1 + \tau\sigma^2\right)^{-(1+a_\eta)}, \tag{B4}$$

which illustrates how the scaling parameter precision τ affects the width of the base distribution for σ.

It is possible to generalize the above 'variable expansion' to a covariance Σ of a multivariate parameter vector ω_n of galaxy intrinsic properties.

Scaling parameter— To complete the specification of the statistical framework when the intrinsic galaxy properties are assumed Gaussian distributed, we derive the conditional posterior for ξ after marginalizing over the galaxy properties ω,

$$\Pr(\xi_n|\mathbf{d}, \sigma_{\eta,n}, m, \tau, \psi) \propto \Pr\left(\xi_n|m, \tau\right) \int d\omega_n \Pr\left(\mathbf{d}_n|\omega_n = (\eta_n |\xi_n|), \psi\right) \Pr(\eta_n|\sigma_{\eta,n}^2), \tag{B5}$$

---

[14] http://thetractor.org/

[15] http://dan.iel.fm/emcee/current/

where again we assume ω_n is univariate for simplicity of presentation. Changing integration variables to $\eta_n = \omega_n/|\xi_n|$, allows us to use the previous importance sampling result (equation (15)) for the marginal likelihood after including the Jacobian for the variable transformation,

$$\Pr(\xi_n|\mathbf{d}, \sigma_{\eta,n}, m, \tau, \psi) \propto \Pr(\xi_n|m, \tau)\frac{1}{N}\sum_{k=1}^{N}|\xi_n|\frac{\Pr\left(\eta_{nk} \equiv \omega_{nk}/\xi_n|\sigma_{\eta,n}^2\right)}{\Pr\left(\omega_{nk}|I_0\right)}. \qquad \text{B6}$$

We use Equation B6 to conditionally update the values of the scaling parameters ξ_n in our MCMC algorithm.

SAMPLING THE DP CLUSTERING PARAMETER

In many applications (including ours) there is no obvious choice for the value of the DP clustering parameter $\mathcal{A}$. Instead, we marginalize over $\mathcal{A}$ with a suitably chosen prior.

Following Dorazio (2009) and Murugiah & Sweeting (2012) we place a conjugate Gamma prior on $\mathcal{A}$ so we can perform Gibbs updates. Dorazio (2009) presents a useful algorithm to determine the hyperparameters (a, b) of the Gamma prior on $\mathcal{A}$ when little information is known about the expected number of clusters in the data (i.e., the number of morphologically distinct galaxy populations).

However, given the extensive existing research into galaxy formation (e.g., the common classifications of 'early' and 'late' type galaxies), we are motivated to place a more informative prior on $\mathcal{A}$. Murugiah & Sweeting (2012) describe several algorithms to determine the parameters (a, b) for the Gamma prior on $\mathcal{A}$ in informative scenarios. We can impose our prior knowledge that there are early- and late-type galaxies by requiring the expected mean value of the number of clusters in the DP to be 2. Using the formula from Antoniak (1974),

$$\mathrm{E}(M|n_{\mathrm{gal}}, \mathcal{A}) \approx \mathcal{A}\ln\left(\frac{n_{\mathrm{gal}} + \mathcal{A}}{\mathcal{A}}\right), \qquad \text{C1}$$

for $n_{\mathrm{gal}} \gg 1$, and where M is the number of clusters or latent classes in the DP. By solving Equation C1 for the value $\bar{\mathcal{A}}$ that gives the expected number of clusters, we set the Gamma prior parameters as,

$$a = \bar{\mathcal{A}}X, \qquad b = X, \qquad \text{C2}$$

so that the prior mean of $\mathcal{A}$ is $a/b = \bar{\mathcal{A}}$ and the prior variance is $a/b^2 = \bar{\mathcal{A}}X$ such that larger X communicates more certainty on the expected number of morphological classes in the galaxy population. Murugiah & Sweeting (2012) present other algorithms to encode prior information on $\mathcal{A}$ that we reserve for future exploration.

Given the hyperprior parameters (a, b), we perform Gibbs updates of $\mathcal{A}$ with the conditional distribution from Escobar & West (1995) (their eq. 13),

$$\mathcal{A}|\zeta, k \sim \pi_\zeta\Gamma(a + k, b - \log(\zeta)) + (1 - \pi_\eta)\Gamma(a + k - 1, b - \log(\zeta)) \qquad \text{C3}$$

where k is the number of latent classes currently assigned to the data, and

$$\pi_\zeta \equiv \frac{a + k - 1}{a + k - 1 + n_{\mathrm{gal}}(b - \log(\zeta))}, \qquad \text{C4}$$

with

$$\zeta|\mathcal{A} \sim \beta\left(\mathcal{A} + 1, n_{\mathrm{gal}}\right), \qquad \text{C5}$$

a variable used only in the Gibbs updates of $\mathcal{A}$, and β denotes the Beta distribution.

GENERATING FAKE POSTERIOR SAMPLES OF GALAXY ELLIPTICITIES

For our 'toy model 2' in Section 3.2.2 we generate fake summary statistics of galaxy pixel data with the following procedure:

1. Specify a value for the true reduced shear g for all galaxies (we assume a constant shear for all galaxies in our numerical examples).

2. For each galaxy $i = 1, \ldots, n_{\mathrm{gal}}$, draw intrinsic ellipticities from a specified distribution $e_i^{\mathrm{int}} \sim \Pr_{e,\mathrm{true}}(\cdot)$.

3. Define the complex ellipticity of the generative model for the pixel data for each galaxy by transforming the intrinsic ellipticity e^{int} with the shear g,

$$\tilde{e}_i = \frac{e_i^{\mathrm{int}} + g}{1 + g^*e_i^{\mathrm{int}}}. \qquad \text{D1}$$

Given the observed ellipticity as a summary statistic of the pixel data, we generate interim posterior samples of the galaxy ellipticity by:

1. Define an interim prior for the model ellipticity $\Pr_{\text{interim}}(\tilde{e}_i)$. This prior may be chosen for computational convenience rather than an accurate description of the data, with the knowledge that we will re-weight the posterior samples later to accomodate a hierarchical PDF on the intrinsic ellipticity e_i^{int}.

2. Define a summary statistic of the pixel data for a galaxy ε_i and a functional form for the likelihood $\Pr(\varepsilon_i|\tilde{e}_i)$ of the summary statistic.

3. Draw samples from the interim posterior,

$$\Pr(\tilde{e}_i|\varepsilon_i) \propto \Pr(\varepsilon_i|\tilde{e}_i)\Pr_{\text{interim}}(\tilde{e}_i).\tag{D2}$$

The samples of $\tilde{e}_i$ contain all the statistical information about galaxy i. For extreme data compression we might limit ourselves to one sample of $\tilde{e}_i$ for each galaxy.

DP SAMPLING ALGORITHM FOR TOY MODEL 2

After marginalizing over the intrinsic ellipticity components, the posterior $\mathcal{H}_n$ for the intrinsic ellipticity variance given a single galaxy n is (assuming weak shear, $g \ll 1$),

$$\begin{aligned}
\mathcal{H}_n(\sigma_\eta^2|\xi_n, \mathbf{d}_n, g) &= G_0(\sigma_\eta^2) \int d^2\eta \, \Pr(\mathbf{d}_n|g, \xi, \eta) \cdot \Pr(\eta|\sigma_\eta^2) \\
&= \frac{(0.5)^{a_\eta}}{\Gamma(a_\eta)} \left(\sigma_\eta^2\right)^{-(1+a_\eta)} \exp\left(-\frac{0.5}{\sigma_\eta^2}\right) \\
&\quad \times \frac{1}{2\pi\left(\sigma_{\text{obs}}^2 + \xi_n^2\sigma_\eta^2\right)} \\
&\quad \times \exp\left(-\frac{1}{2}\frac{(d_{n,1}-g_1)^2 + (d_{n,2}-g_2)^2}{\sigma_{\text{obs}}^2 + \xi_n^2\sigma_\eta^2}\right).
\end{aligned}\tag{E1}$$

This expression is pedagogical only. In practice, we do not have $\mathbf{d}_n$ values. Instead we have importance samples of $e^{\text{int}} \equiv |\xi|\eta$ with which we perform the above integral via Monte Carlo.

Following from Equation (21) and Neal (2000), the conditional posterior for $\sigma_{\eta,n}^2$ for galaxy n is,

$$\sigma_{\eta,n}^2|\sigma_{\eta,-n}^2, \mathbf{d}_n \sim \sum_{m\neq n} q_{n,m}\delta_{\text{D}}\left(\sigma_{\eta,m}^2\right) + r_n\mathcal{H}_n\tag{E2}$$

where $\delta_{\text{D}}(\sigma_{\eta,m}^2)$ is a Dirac delta function indicating that $\sigma_{\eta,n}^2$ for galaxy n takes the same value as that for galaxy m. This is how clustering occurs in the Gibbs sampling. The coefficient of the delta functions $q_{n,m}$ is defined to be the marginal likelihood for galaxy n given the parameter $\sigma_{\eta,m}^2$ currently assigned to galaxy m,

$$q_{n,m} = b\int d^2\eta \, \Pr(\mathbf{d}_n|\eta, \xi_m, g) \cdot \Pr(\eta|\sigma_{\eta,m}^2),\tag{E3}$$

where b is a constant defined below. If all the $q_{n,m}\forall m \neq n$ are sufficiently small, meaning that no parameters σ^2 currently assigned to other galaxies are good descriptions of observation n, then we draw a new value for $\sigma_{\eta,n}^2$ from H_n. The coefficient that defines the probability for drawing new values relative to the coefficients $q_{n,m}$ is the likelihood for galaxy n marginalized over both the intrinsic ellipticity components and the hyperprior parameter σ_η^2,

$$r_n = b\,\mathcal{A}\int d\sigma_\eta^2 \int d^2\eta \, \Pr(\mathbf{d}_n|\eta, \xi_n, g) \cdot \Pr(\eta|\sigma_\eta^2) \cdot G_0(\sigma_\eta^2|a_\eta).\tag{E4}$$

Note that r_n is proportional to $\mathcal{A}$ so that larger $\mathcal{A}$ values mean less clustering of $\sigma_{\eta,n}^2$ values for each galaxy.

The parameter b is defined so that

$$\sum_{m\neq n} q_{n,m} + r_n = 1.\tag{E5}$$

Given the importance samples for e^{int}, we compute $q_{n,m}$ as,

$$\frac{1}{b}q_{n,m} = \frac{1}{K}\sum_{k=1}^{K}\frac{\Pr(e_{kn}^{\text{int}}/\xi_n|\sigma_{\eta,m}^2)}{\Pr_{\text{interim}}(e_{kn}^{\text{int}})}\tag{E6}$$

Similarly, we compute r_n by first integrating analytically over σ_η^2 and then summing over the e^{int} samples,

$$\frac{1}{b}r_i = \frac{1}{K}\sum_{k=1}^{K}\frac{\Pr_{\text{marg}}(e_{kn}^{\text{int}}/\xi_n|a_\eta)}{\Pr_{\text{interim}}(e_{kn}^{\text{int}})},\tag{E7}$$

where, assuming the parameter $m = 0$ as in Equation B4,

$$\mathrm{Pr}_{\mathrm{marg}}\left(\eta|a_\eta\right) \equiv \frac{0.5^{a_\eta}}{2\pi}\frac{\Gamma(1+a_\eta)}{\Gamma(a_\eta)}\left(\frac{1}{2}\left(1+|\eta|^2\right)\right)^{-(1+a_\eta)} \tag{E8}$$

and a_η is a parameter in the inverse Gamma base distribution G_0 as defined in Equation B3.

COMPUTATIONAL IMPLEMENTATION

In Figure 9 we show the ordering as well as inputs and outputs of the distinct computations in an implementation of our statistical framework for cosmological parameter inference from cosmic shear. To aide in the conceptual organization of the computations, we introduce an analogy to farming by considering the 'harvesting' and processing of information from pixel data in a galaxy imaging survey.

We assume as inputs CCD imaging data of a field of galaxies, perhaps with several exposures or 'multiple epochs'. The first step is to identify distinct sources (i.e., galaxies and stars) in the images and select small regions of pixel data around each source (called 'cutouts'). In this paper, we do not specify an algorithm for identifying distinct sources. In principle we should specify a probabilistic algorithm, particularly for the ambiguous cases of separating stars and galaxies (Soumagnac et al. 2013) or blending of distinct sources by the PSF (Dawson et al. 2014). A probabilistic source extraction algorithm will be particularly important for combining imaging at different resolutions and wavelengths (Dawson & Schneider 2014; Jain et al. 2015). But, for this paper we assume that distinct source cutouts are available by some other means.

Each galaxy image cutout is supplied as input to the REAPER algorithm to draw posterior samples of galaxy image model parameters ω assuming an interim prior that is statistically independent for each galaxy. The model parameters for the PSF Π can also be sampled with an interim prior within the REAPER. The output of the REAPER algorithm includes MCMC posterior samples for galaxy and PSF models for every distinct galaxy in the imaging data. Allowing for $N \sim 10^2$ parameters per galaxy to describe features of the galaxy image and PSF, $M \sim 10$ saved interim posterior samples per galaxy, and $P \sim 10^6$ galaxies in a single pointing of the LSST, we estimate a total of $\sim 10^9$ floating point numbers as outputs from the REAPER. This would require ~ 10Gb of disk space per field. This modest disk storage requirement is enabled by the assumption that we will only need to store a small number (~ 10) interim samples per object. There are of order 10^3 pointings in the LSST survey, yielding a total REAPER output of ~ 1Tb.

The hierarchical model for the galaxy parameter distributions is implemented in the THRESHER algorithm. The THRESHER takes as input the MCMC samples from REAPER for all galaxies in a region of the sky. Using importance sampling as in Equation (15) and Equation (16), we marginalize the galaxy model parameters to infer simultaneous constraints on the parameters of the distribution of galaxy parameters and the lens potential ψ or, alternatively, the shear components. The output of the THRESHER algorithm includes posterior samples of α_n, ξ_n, and ψ. These samples may require much larger disk storage because of the large number of galaxies indexed by $n \sim 10^9$. The lens potential ψ MCMC samples may be considered to be represented on a 2D grid as interpolated from the observed galaxy positions on the sky. To resolve several Gpc scales down to halo scales $\sim$Mpc, the number of grid points for representing ψ must be $P \sim (3 \times 10^3)^2 \approx 10^7$. With again $M \sim 10$ interim samples saved from the THRESHER, we are required to store $(10^9 \times 10^2 + 10^7 \text{ parameters}) \times (10 \text{ samples}) \approx 10^{12}$ floating point numbers from the THRESHER, or ~ 10 Tb.

Finally, the THRESHER output MCMC samples of ψ are supplied as inputs to the WINNOWER algorithm for inference of cosmological parameters via importance sampling as in Equation (17). The WINNOWER may allow combining ψ_s samples output from separate THRESHER runs with different source galaxy samples s. The source samples must be combined in the WINNOWER because of the common lensing mass for all source samples.
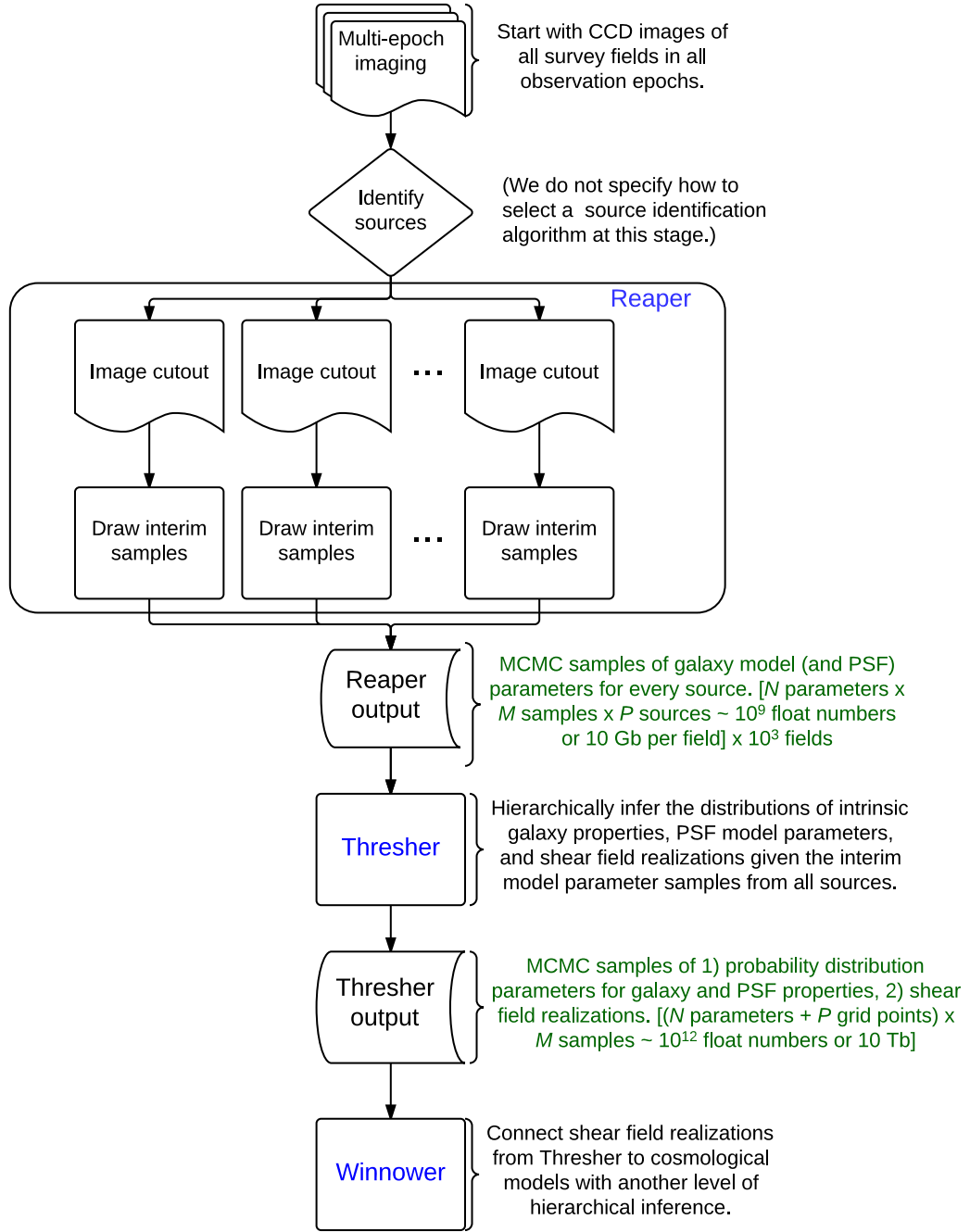
Fig. 9.— The computational steps in the implementation of our hierarchical model for shear inference.